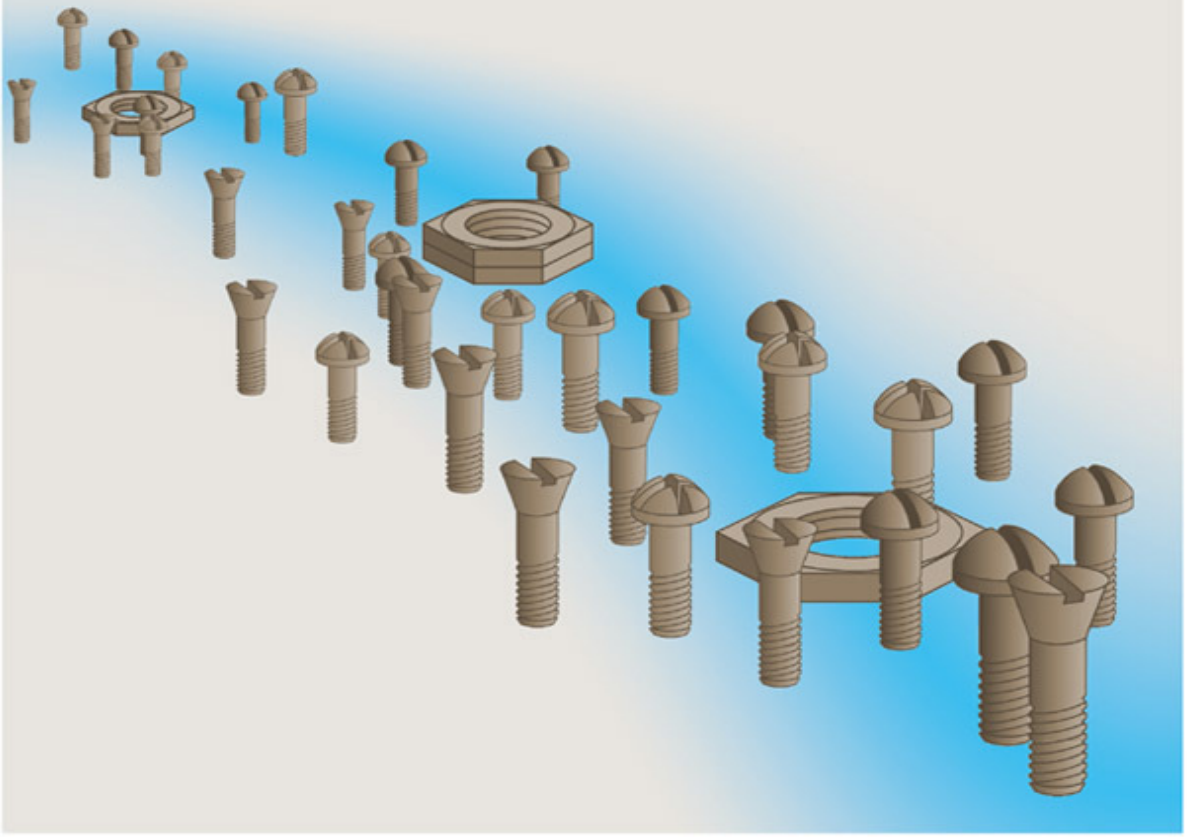
Sociología

Jon Elster

La explicación del comportamiento social

Más tuercas y tornillos para las ciencias sociales



gedisa editorial

Otros títulos de interés:

Introducción a la sociología
Theodor W. Adorno

Trabajo, consumismo y nuevos pobres
Zigmunt Bauman

El racismo: una introducción
Michel Wieviorka

Cultura y explosión
Lo previsible y lo imprevisible en los procesos de cambio social
Yuri M. Lotman

Sociología simétrica
Miquel Doménech y Francisco J. Tirado
(Coords.)

Sociología fundamental
Norbert Elías

La esperanza de Pandora
Ensayo sobre la realidad de los estudios de la ciencia
Bruno Latour

Historia de la idea de progreso
Robert Nisbet

Grupos, organizaciones e instituciones
La transformación de la burocracia
Georges Lapassade

La cultura del control
Crimen y orden social en la sociedad contemporánea
David Garland

Jon Elster

LA EXPLICACIÓN DEL COMPORTAMIENTO SOCIAL

Serie CLA•DE•MA
Filosofía

Obras de
JON ELSTER
publicadas por
Editorial Gedisa

Ulises desatado

La democracia deliberativa

La ética de las decisiones médicas

"Egonomics"

Psicología política

Justicia local
De qué modo las instituciones distribuyen bienes escasos y cargas necesarias

Lógica y sociedad

Juicios salomónicos
Las limitaciones de la racionalidad como principio de decisión

El cemento de la sociedad
Las paradojas del orden social

Tuercas y tornillos
Una introducción a los conceptos básicos de las ciencias sociales

El cambio tecnológico
Investigaciones sobre la racionalidad y la transformación social

La explicación del comportamiento social
Más tuercas y tornillos para las ciencias sociales

LA EXPLICACIÓN DEL COMPORTAMIENTO SOCIAL

Jon Elster

Traducción de Horacio Pons

gedisa
editorial

Textos del original inglés:
Explaining Social Behaviour. More Nuts and Bolts for the Social Sciences

© 2007, Cambridge University Press

Creative Commons

Traducción: Horacio Pons
Diseño de cubierta: Kaffa

Primera edición, septiembre de 2010, Barcelona

Derechos reservados para todas las ediciones en castellano

© *by* Editorial Gedisa S. A.
Avenida del Tibidabo, 12, 3º
Tel. 34 93 253 09 04
Fax 34 93 253 09 05
08022 - Barcelona, España
gedisa@gedisa.com
www.gedisa.com

ISBN 978-84-9784-251-8
Depósito legal: B. 42.454 - 2010

Impreso por Sagrafic

Impreso en España
Printed in Spain

Esta obra ha sido publicada con una subvención de la Dirección General del Libro, Archivos y Bibliotecas del Ministerio de Cultura, para su préstamo público en Bibliotecas Públicas, de acuerdo con lo previsto en el artículo 37.2 de la Ley de Propiedad Intelectual

Para Jonathan y Juana

Índice

Prefacio
Introducción

Primera parte: Explicación y mecanismos

1 Explicación
2 Mecanismos
3 Interpretación

Segunda parte: La mente

4 Motivaciones
5 Egoísmo y altruismo
6 Miopía y previsión
7 Creencias
8 Emociones

Tercera parte: Acción

9 Deseos y oportunidades
10 Personas y situaciones
11 Elección racional
12 Racionalidad y comportamiento
13 Respuestas a la irracionalidad
14 Algunas implicaciones para la interpretación textual

Cuarta parte: Lecciones de las ciencias naturales

15 Fisiología y neurociencia
16 Explicación por las consecuencias y selección natural
17 Selección y comportamiento humano

Quinta parte: Interacción

18 Consecuencias indeliberadas
19 Interacción estratégica
20 Juegos y comportamiento
21 Confianza
22 Normas sociales
23 Formación colectiva de creencias
24 Acción colectiva
25 Toma colectiva de decisiones
26 Organizaciones e instituciones

Conclusión: ¿Es posible la ciencia social?

Proverbios y máximas citados

Prefacio

Este libro comenzó siendo una revisión de un volumen que publiqué en 1989, *Nuts and Bolts for the Social Sciences*.* Y terminó por adquirir un carácter muy diferente y más ambicioso. Abarca una variedad mucho más grande de temas, con una atención considerablemente más importante al detalle y un espíritu diferente. Si bien nueve capítulos conservan los títulos del libro anterior, sólo el Capítulo 9 y el Capítulo 24 se mantienen sustancialmente sin cambios.

Aunque de alcance general, este volumen no es un tratado. Es, a la vez, más y menos. Es una presentación elemental, informal y personal de ideas que tienen, creo, un notable potencial para esclarecer el comportamiento** social. Utilizo numerosos ejemplos, muchos de ellos anecdóticos o literarios y otros tomados de estudios más sistemáticos. El uso muy ocasional del álgebra no va más allá del nivel de la escuela secundaria. Al mismo tiempo, el libro tiene una inclinación metodológica y filosófica que no es habitual en las presentaciones de carácter introductorio. Hay en él un esfuerzo por situar las ciencias sociales dentro del ámbito más general de las ciencias, tanto las naturales como las humanidades. Hay también un afán por mantener en el lector un alerta constante acerca del hecho de que los principios generales de la explicación científica ponen límites a la construcción de teorías con pretensiones explicativas.

El estilo de las notas bibliográficas incluidas al final de cada capítulo refleja el avance de Internet, sobre todo de Wikipedia, google.com y scholar.google.com. Como los lectores pueden encontrar la mayoría de las referencias relevantes en cuestión de minutos, he omitido las fuentes de muchos de los enunciados y descubrimientos del texto. Procuré en cambio indicarles libros de consulta importantes, algunos clásicos modernos,

volúmenes y artículos que son el origen de ideas que tal vez sea difícil rastrear en Internet, y autores de quienes he tomado tanto que el no mencionarlos justificaría un juego de palabras con mi apellido (en alemán, *Elster* significa urraca).

Aunque el texto principal contiene pocas referencias a estudiosos contemporáneos, me refiero *in extenso* a Aristóteles, Séneca, Montaigne, La Rochefoucauld, Samuel Johnson, H. C. Andersen, Stendhal, Tocqueville, Proust y otros escritores clásicos que siguen siendo fuentes literalmente inagotables de hipótesis causales. Ignoraríamos muchas ideas perspicaces si hiciéramos caso omiso de los mecanismos sugeridos por la filosofía, la ficción, el teatro y la poesía. Si pasamos por alto veinticinco siglos de reflexión sobre la mente, la acción y la interacción para concentrarnos en los cien o diez últimos años, lo hacemos a nuestro riesgo y en desmedro de nuestra posición. Cito a esos autores no tanto para apelar a su autoridad como para justificar el argumento de que más vale que nuestras lecturas sean amplias y no restringidas. En contraste directo con la inexorable profesionalización de la ciencia social (sobre todo la estadounidense), que no promueve en los estudiantes el aprendizaje de idiomas extranjeros y la lectura de viejos libros, el presente volumen es un extenso alegato por un enfoque más abarcativo del estudio de la sociedad.

* * *

En la preparación del manuscrito conté con la asistencia y los comentarios de muchas personas. Debo agradecer ante todo a mis estudiantes de la Universidad de Columbia por sus incisivas preguntas y observaciones en el curso donde presenté por primera vez el material que terminó por convertirse en este libro. Las sugerencias de Pablo Kalmanovitz fueron de particular utilidad. En Collioure, Aanund Hylland y Ole-Jørgen Skog dedicaron tres días a discutir conmigo un borrador del libro entero. En Oslo, el propio Hylland, Karl O. Moene y John Roemer continuaron la discusión durante un día y medio. Sus comentarios no sólo me salvaron de muchos (¡muchos!) errores, sino que también sugirieron la manera de complementar y consolidar la exposición. Estoy agradecido a Roemer, en especial, por instarme a escribir una conclusión. Recibí comentarios escritos sobre todo el manuscrito de Diego Gambetta, Raj Saah y un reseñador anónimo. Los comentarios de Gambetta fueron particularmente detallados y

útiles. Mantuve fructíferas conversaciones con Walter Mischel sobre las ideas que son en gran parte de su propia cosecha, presentadas en el Capítulo 10. También recibí valiosos comentarios escritos de George Ainslie sobre las ideas –muchas de ellas planteadas por él– expuestas en la primera parte del libro. Bernard Manin hizo constructivas observaciones sobre el Capítulo 25. Robyn Dawes propuso agudos comentarios sobre el Capítulo 7 y el Capítulo 12. Para terminar, a lo largo de los últimos años he presentado versiones preliminares de capítulos del libro a los miembros del «grupo de los lunes» que, desde 1995, se ha reunido semanalmente en la ciudad de Nueva York todos los otoños y, de manera más ocasional, en la primavera: John Ferejohn, Raquel Fernández, Russell Hardin, Stephen Holmes, Steven Lukes, Bernard Manin, Pasquale Pasquino, Adam Przeworski y John Roemer. Agradezco a todos ellos sus objeciones amistosas y constructivas.

Dedico el libro a Jonathan y Joanna Cole; ellos saben por qué.

* * *

Cito los *Ensayos* de Montaigne de la traducción de M. Screech (Londres, Penguin, 1971); a Proust, de la nueva traducción editada por C. Prendergast (Londres, Penguin, 2003); los *Pensamientos* de Pascal, de la traducción de A. J. Krailsheimer (Londres, Penguin, 1995); las *Máximas* de La Rochefoucauld, de la traducción de L. Tancock (Londres, Penguin, 1981); *Los caracteres* de La Bruyère, de la traducción de H. van Laun (Nueva York, Scribner, 1885); *Del amor*, de Stendhal, de la traducción de G. Sale, S. Sale y J. Stewart (Londres, Penguin, 1975), y *La democracia en América*, de Tocqueville, de la nueva traducción de A. Goldhammer (Nueva York, Library of America, 2004). Otras traducciones del francés son mías.*

Notas al pie

* Hay traducción en castellano: *Tuercas y tornillos. Una introducción a los conceptos básicos de las ciencias sociales*, Barcelona, Gedisa, 2003, 3ª ed. [T.]

** En este libro se han tomado como sinónimos las palabras «comportamiento» y «conducta». [E.]

* Menciono a continuación algunas de las versiones en castellano que he utilizado en esta traducción: Michel de Montaigne, *Ensayos*, tres volúmenes, traducción de María Dolores Picazo y Almudena Montojo, Barcelona, Altaya, 1994; Marcel Proust, *En busca del tiempo perdido*, siete volúmenes, traducción de Pedro Salinas, José María Quiroga Plà y Consuelo Berges, Madrid, Alianza, 1966-1969; Blaise Pascal, *Pensamientos*, traducción de Juan Domínguez Berrueta, Buenos Aires, Orbis, 1984; Stendhal, *Del amor*, traducción de Consuelo Berges, Madrid, Alianza, 1968, y Alexis de Tocqueville, *La democracia en América*, traducción de Luis R. Cuéllar, México, Fondo de Cultura Económica, 1957. [T.]

Introducción

Este libro se ocupa de la explicación del comportamiento social. En la primera parte, expongo mi concepción de la explicación, y en las cuatro partes restantes construyo una caja de herramientas de conceptos y mecanismos que se aplican a casos particulares. No hace falta decir que el libro no aspira a ser exhaustivo. En vez de tratar de señalar las lagunas, que serán evidentes, me gustaría comenzar por presentar una muestra de los enigmas que, a mi entender, pueden ser dilucidados por el enfoque que adopto. En la conclusión, vuelvo a esos mismos enigmas con breves referencias a las explicaciones que he citado en capítulos anteriores.

Los ejemplos y las explicaciones deben considerarse con dos salvedades en mente. En primer lugar, no pretendo que todos los *explananda* son hechos bien establecidos. En una explicación concreta, éste es desde luego un primer paso crucial: no tiene sentido intentar explicar lo que no existe. Sin embargo, cuando el objetivo es construir una caja de herramientas, uno puede ser menos riguroso. En segundo lugar, aun para los *explananda* cuya existencia está bien documentada, no afirmo que las explicaciones que cito sean las correctas. Sólo sostengo que cumplen una condición mínima de la explicación: implican lógicamente los *explananda*. Los enigmas y las explicaciones pretenden mostrar que «si pasa este tipo de cosas, aquí está el tipo de mecanismo que podría explicarlo», así como que «si actúa este mecanismo, aquí está el tipo de cosas que puede producir». Hechas estas salvedades, expongo a continuación los enigmas, ordenados de manera un tanto arbitraria (puesto que muchos de ellos podrían incluirse en varias categorías) en relación con las cuatro partes sustantivas del libro.[1]

PRIMERA PARTE: LA MENTE

- ¿Por qué algunos jugadores creen que cuando el rojo ha salido cinco veces seguidas, tiene más probabilidades que el negro de volver a salir la vez siguiente?
- ¿Por qué otros jugadores creen que cuando el rojo ha salido cinco veces seguidas, el negro tiene más probabilidades de salir la vez siguiente?
- ¿Por qué las preferencias cambian a veces con el mero paso del tiempo?
- ¿Por qué tanta gente que parece creer en la vida después de la muerte quiere que ésta llegue lo más tarde posible?
- ¿Por qué la gente es reacia a reconocerse y reconocer ante los demás, que es envidiosa?
- ¿Por qué la gente es reacia a reconocerse y reconocer ante los demás, que es ignorante?
- ¿Por qué, entre los conversos del siglo XVI al calvinismo, la creencia de que la gente estaba predestinada o bien al cielo o bien al infierno inducía más paz espiritual que la creencia de que era posible alcanzar la salvación mediante las buenas obras?
- ¿Por qué es verdad (a veces) que «quien ha ofendido, no puede perdonar»?
- ¿Por qué en algunas culturas la vergüenza es más importante que la culpa?
- ¿Por qué la victoria francesa en la Copa del Mundo de fútbol de 1998 generó tanta alegría en el país, y por qué el hecho de que el equipo francés no pasara de la primera fase en 2002 causó tanto desazón?
- ¿Por qué las mujeres a menudo sienten vergüenza después de haber sido violadas?
- ¿Por qué los rituales humillantes de iniciación generan más y no menos lealtad al grupo en el cual uno es iniciado?

SEGUNDA PARTE: ACCIÓN

- ¿Por qué hoy son más numerosos que hace veinte años los espectáculos de Broadway que suscitan ovaciones de pie del público?

- ¿Por qué los castigos incrementan, en vez de disminuir, la frecuencia del comportamiento al que apuntan?
- ¿Por qué la gente es reacia a romper reglas autoimpuestas, aun en los casos en que tiene poco sentido respetarlas?
- ¿Por qué el modelo de venganza de alguna gente es «dos ojos por uno» y no «ojo por ojo»?
- ¿Por qué el rendimiento a largo plazo de las acciones es muy superior al de los bonos (es decir, por qué el valor de las acciones no sube hasta igualar los rendimientos)?
- ¿Por qué los índices de suicidio disminuyen cuando los medicamentos peligrosos se venden en blísteres y no en frascos?
- ¿Por qué ninguno de treinta y ocho espectadores llamó a la policía cuando Kitty Genovese fue asesinada a golpes?
- ¿Por qué algunos individuos escondieron o rescataron a judíos bajo el régimen nazi?
- ¿Por qué el presidente Chirac convocó a elecciones anticipadas en 1997, con el único resultado de perder su mayoría parlamentaria?
- ¿Por qué algunos padres que se divorcian están dispuestos a compartir la custodia de los hijos cuando la solución que prefieren es la custodia exclusiva, que probablemente obtendrían si litigaran?
- ¿Por qué los pobres son menos propensos a emigrar?
- ¿Por qué algunas personas depositan ahorros en cuentas navideñas que no pagan intereses y no permiten retiros antes de Navidad?
- ¿Por qué la gente encara proyectos, como la construcción del avión Concorde, que tienen un valor esperado negativo?
- ¿Por qué, en la «justicia transicional» (situación en que los agentes de un régimen autocrático son sometidos a juicio luego de la transición a la democracia), quienes son juzgados inmediatamente después de la transición reciben sentencias más severas que quienes son juzgados más adelante?
- ¿Por qué, en la obra de Shakespeare, Hamlet posterga la venganza hasta el último acto?

TERCERA PARTE: LECCIONES DE LAS CIENCIAS NATURALES

- ¿Por qué es mucho más probable que los padres maten a hijos adoptados o hijastros que a sus hijos biológicos?

- ¿Por qué es tan escaso el incesto entre hermanos, si se tienen en cuenta las tentaciones y oportunidades?
- ¿Por qué la gente invierte su dinero en proyectos emprendidos por otros agentes, aun cuando éstos tengan la libertad de apropiarse de todos los beneficios?
- ¿Por qué, al vengarse, la gente acepta pagar algún coste material, sin recibir beneficios de la misma especie?
- ¿Por qué la gente se apresura a sacar conclusiones que las pruebas existentes no justifican?

CUARTA PARTE: INTERACCIÓN

- ¿Por qué los adherentes de un partido socialista a veces votan a los comunistas y de ese modo impiden la victoria de su partido?
- ¿Por qué algunos países que se han independizado recientemente adoptan como idioma oficial el de su anterior opresor imperialista?
- ¿Por qué los puestos de helados de la playa suelen estar unos al lado de otros, cuando los clientes estarían mejor atendidos y a los vendedores no les iría peor, si estuvieran más separados entre sí?
- ¿Por qué un individuo vota en elecciones cuando es virtualmente cierto que su voto no tendrá influencia en el resultado?
- ¿Por qué los individuos económicamente exitosos de las sociedades occidentales modernas suelen ser más delgados que la persona común y corriente?
- ¿Por qué la gente se abstiene de hacer transacciones que podrían mejorar la situación de todos, por ejemplo, cuando evita preguntar a la primera persona de la fila en una parada de autobús si está dispuesta a vender su lugar?
- ¿Por qué el presidente Nixon trató de mostrarse ante los soviéticos como si fuera propenso a un comportamiento irracional?
- ¿Por qué los jefes militares a veces queman los puentes (o sus propias naves)?
- ¿Por qué la gente suele atribuir gran importancia a cuestiones de etiqueta intrínsecamente insignificantes?
- ¿Por qué los pasajeros dan propina al conductor de un taxi y los comensales hacen lo mismo con un camarero, aun cuando se

encuentren en una ciudad extranjera a la que no tienen intención de volver?
- ¿Por qué las empresas hacen extensos inventarios aunque no prevean ninguna interrupción de la producción?
- ¿Por qué, en un grupo de estudiantes, cada uno piensa que los otros han entendido mejor que él un texto oscuro?
- ¿Por qué en muchas asambleas políticas se vota pasando lista?
- ¿Por qué el intercambio de favores es más frecuente en las legislaturas ordinarias que en las asambleas constituyentes?

Proporcionaré las explicaciones sugeridas de estos fenómenos en distintos lugares del libro, y las resumiré brevemente en la conclusión. Aquí sólo quiero hacer una observación general acerca de dos tipos de explicación cuya utilidad es *improbable*. Como los lectores advertirán ya en el Capítulo 1, con varios recordatorios a lo largo del camino, uno de los objetivos del libro es inculcar escepticismo con respecto a dos líneas comunes de razonamiento. En primer lugar, las ciencias sociales no pueden, con muy pocas excepciones, apoyarse en la explicación funcional, que explica acciones o patrones de comportamiento mediante el recurso a sus consecuencias y no a sus causas. ¿Las normas relacionadas con la propina existen porque es más eficaz hacer que los clientes, y no los propietarios, controlen el comportamiento de los camareros? Creo que no. En segundo lugar, hoy me parece que la teoría de la elección racional tiene menos capacidad explicativa de lo que yo suponía. ¿Actúan las personas comunes sobre la base de los cálculos que llenan tantas páginas de anexos matemáticos en las principales revistas especializadas? Creo que no.

Al menos en tres aspectos, la teoría de la elección racional es, no obstante, un valioso elemento de la caja de herramientas. Si se la entiende en un sentido cualitativo práctico, es capaz de explicar gran parte del comportamiento cotidiano. Y aun cuando no explique mucho, puede tener un enorme valor conceptual. La teoría de los juegos, en particular, ha esclarecido la estructura de la interacción social de una manera que supera con mucho las intuiciones que al respecto se alcanzaron en siglos anteriores. Por último, los seres humanos *quieren* ser racionales. El deseo de contar con suficientes razones de nuestra conducta, y no ser meros juguetes de fuerzas psíquicas que actúan «a nuestras espaldas», representa

una fuerza permanente contrapuesta a los numerosos mecanismos generadores de irracionalidad que examino en este libro.

Si bien soy crítico de muchas explicaciones basadas en la elección racional, creo que el concepto de *elección* es fundamental. En el libro, considero varias alternativas a la explicación fundada en la elección y llego a la conclusión de que, aunque a veces quizá sean de utilidad para complementar ese enfoque, no pueden reemplazarla. El hecho de que la gente actúe bajo diferentes *coacciones*, por ejemplo, puede explicar con frecuencia una gran cantidad de variaciones en el comportamiento. Además, en algunos casos es posible argumentar que la *selección de agentes*, y no la *elección por parte de agentes*, es responsable del comportamiento que observamos. En líneas generales, sin embargo, creo que el factor subjetivo de la elección tiene mayor capacidad explicativa que los factores objetivos de las coacciones y la selección. Como es obvio, ésta es una intuición para la que no hay ninguna demostración rigurosa. Lo cual no impide que, de todos modos, los científicos sociales deban dar cabida a la totalidad de los factores en su caja de herramientas.

Nota

[1] Aunque la lista se superpone un poco con una serie de enigmas presentados en el Capítulo 12 como desafíos a la teoría de la elección racional, su propósito no es polémico; sólo aspira a incitar la curiosidad del lector.

PRIMERA PARTE

Explicación y mecanismos

Este libro se apoya en una concepción específica de la explicación en las ciencias sociales. Aunque no es en lo primordial una obra de filosofía de la ciencia social, sostiene y apela a ciertas ideas metodológicas sobre el modo de explicar los fenómenos sociales. En los primeros tres capítulos, esas ideas se exponen en forma explícita. En el resto del volumen, forman parte sobre todo del trasfondo implícito, aunque de vez en cuando, especialmente del Capítulo 14 al Capítulo 17 y en la conclusión, vuelven a ocupar el centro del escenario.

Mi argumento es que todas las explicaciones son causales. Explicar un fenómeno (un *explanandum*) es citar un fenómeno anterior (el *explanans*) que lo ha causado. Al abogar por la explicación causal, no pretendo excluir la posibilidad de una explicación intencional del comportamiento. Las intenciones pueden actuar como causas. Una variedad particular de explicación intencional es la *explicación basada en la elección racional*, que será objeto de extensos análisis en capítulos siguientes. Sin embargo, muchas explicaciones intencionales se fundan en el supuesto de que los agentes son, de una manera u otra, *irracionales*. En sí misma, la irracionalidad no es más que una idea negativa o residual: todo lo que no es racional. Para que tenga algún valor explicativo, es menester apelar a formas específicas de irracionalidad con implicaciones específicas para el comportamiento. En el Capítulo 12, por ejemplo, enumero e ilustro once mecanismos que pueden generar una conducta irracional.

En ocasiones, los científicos explican los fenómenos por sus *consecuencias* y no por sus causas. Tal vez digan, por ejemplo, que las rivalidades sangrientas se explican por el hecho de que reducen la población a niveles sustentables. Esta idea podría parecer una imposibilidad metafísica: ¿cómo es posible explicar la existencia u ocurrencia de algo en un momento por medio de otra cosa que todavía no tiene entidad? Como veremos, el problema puede reformularse para convertir la explicación por las consecuencias en un concepto valedero. En las ciencias biológicas, la explicación evolucionista es un ejemplo de ello. En las ciencias sociales, sin embargo, los ejemplos eficaces de este tipo de explicación son pocos y están alejados entre sí. El ejemplo de la rivalidad de sangre no es, definitivamente, uno de ellos.

Las ciencias naturales, en especial la física y la química, proponen *explicaciones basadas en leyes*; las leyes son proposiciones generales que nos permiten inferir la verdad de un enunciado de la verdad de otro anterior

en el tiempo. De tal modo, cuando conocemos las posiciones y la velocidad de los planetas en un momento determinado, las leyes del movimiento planetario nos permiten deducir y predecir sus posiciones en cualquier momento ulterior. Este tipo de explicación es *determinista*: dados los antecedentes, sólo es posible un consecuente. En materia de este tipo de explicaciones basadas en leyes, las ciencias sociales ofrecen poco y nada. La relación entre *explanans* y *explanandum* no es de uno a uno o de muchos a uno, sino de uno a muchos o de muchos a muchos. Numerosos especialistas en ciencias sociales tratan de modelizar esa relación mediante el uso de métodos *estadísticos*. Sin embargo, las explicaciones estadísticas son incompletas de por sí, dado que, en última instancia, tienen que fundarse en intuiciones sobre *mecanismos* causales verosímiles.

Capítulo 1
Explicación

Explicación: general

La principal tarea de las ciencias sociales es explicar los fenómenos sociales. No es la única sino la más importante, la tarea a la cual las demás están subordinadas o de la cual dependen. El tipo básico de *explanandum* es un *suceso*. Explicarlo es dar razón de por qué sucedió, mencionando como causa un *suceso anterior*. Así, podemos explicar la victoria de Ronald Reagan en las elecciones presidenciales de 1980 por el fracasado intento de Jimmy Carter de rescatar a los rehenes estadounidenses en Irán.[1] Podríamos asimismo explicar el estallido de la Segunda Guerra Mundial con referencia a un número cualquiera de otros sucesos, desde el Pacto de Múnich hasta la firma del Tratado de Versalles. Si bien en ambos casos la estructura fina de la explicación causal será obviamente más compleja, los ejemplos citados encarnan el patrón de explicación basada en *dos sucesos* ocurridos uno tras otro [*suceso-suceso*]. En una tradición que tiene su origen en David Hume, suele hacerse referencia a él como el modelo de la «bola de billar». Un suceso, la bola A que golpea la bola B, es la causa de (y con ello explica) otro suceso, a saber, la puesta en movimiento de la bola B.

Quienes están familiarizados con el tipo característico de explicación de las ciencias sociales tal vez no reconozcan ese patrón o no lo consideren como privilegiado. De una manera u otra, los especialistas en ciencias sociales tienden a hacer más hincapié en los *hechos*, o estados de cosas, que en los sucesos. La oración «A las nueve de la mañana la ruta estaba

resbaladiza» enuncia un hecho. La oración «A las nueve de la mañana el automóvil se salió de la ruta» enuncia un suceso. Como lo sugiere este ejemplo simple, se podría recurrir al modelo de *hecho-suceso* para explicar un accidente.[2] A la inversa, podría apelarse al modelo de *suceso-hecho* para explicar un estado determinado de cosas, como ocurre cuando afirmamos que el ataque al World Trade Center en 2001 explica el miedo generalizado de muchos estadounidenses. Para terminar, las explicaciones convencionales de las ciencias sociales tienen con frecuencia un patrón de *hecho-hecho*. Tomemos un ejemplo al azar: se ha sostenido que el nivel de educación de las mujeres explica el ingreso per cápita en el mundo en vías de desarrollo.

Consideremos la explicación de un hecho en particular, el que el 65% de los estadounidenses apoyen o digan apoyar la pena de muerte.[3] En principio, este problema puede reformularse en términos de sucesos: ¿cómo *han llegado* estos estadounidenses *a apoyar* la pena de muerte? ¿Cuáles fueron los sucesos formativos, interacciones con los padres, los pares o los docentes, que provocaron la aparición de esa actitud? En la práctica, los especialistas en ciencias sociales no suelen interesarse en la cuestión. En vez de procurar explicar una estadística en bruto de este tipo, quieren entender los *cambios* en las actitudes a lo largo del tiempo o las *diferencias* de actitudes a través de las poblaciones. La razón es, tal vez, que el hecho en bruto no les parece muy informativo. Si se pregunta si el 65% es mucho o poco, la réplica obvia es: «¿en comparación con qué?» En comparación con las actitudes de los estadounidenses alrededor de 1990, cuando aproximadamente el 80% estaba a favor de la pena capital, es un número bajo. En comparación con las actitudes preponderantes en algunos países europeos, es un número alto.

Los estudios *longitudinales* consideran las variaciones a lo largo del tiempo en la variable dependiente. Los estudios *transversales* consideran las variaciones a través de poblaciones. En uno y otro caso, el *explanandum* se transforma. En vez de tratar de explicar el fenómeno «en y por sí mismo», tratamos de explicar cómo varía en el tiempo o el espacio. El éxito de una explicación se mide en parte por la proporción de la «varianza» (una medida técnica de variación) que puede explicar.[4] Un éxito completo explicaría todas las variaciones observadas. En un estudio transversal nacional podríamos comprobar, por ejemplo, que el porcentaje de individuos favorables a la pena de muerte es estrictamente proporcional a la

cantidad de homicidios cada cien mil habitantes. Si bien este descubrimiento no daría *ninguna* explicación de los números absolutos, sí ofrecería una explicación *perfecta* de la diferencia entre ellos.[5] En la práctica, desde luego, el éxito perfecto nunca se alcanza, pero da igual. Las explicaciones de la varianza no dicen nada sobre el *explanandum* «en y por sí mismo».

Puede tomarse un ejemplo del estudio del comportamiento de los votantes. Como veremos más adelante (en el Capítulo 12), no resulta claro por qué los ciudadanos se molestan en votar en las elecciones nacionales, cuando es moralmente indudable que un solo voto no significará diferencia alguna. Pese a ello, un porcentaje sustancial del electorado concurre a votar el día de los comicios. ¿Por qué se toman la molestia de hacerlo?

En vez de intentar resolver este misterio, los sociólogos empíricos suelen abordar una cuestión diferente: ¿por qué varía la concurrencia en distintas elecciones? Una hipótesis subraya la probabilidad de que los votantes tiendan a abstenerse de concurrir cuando el tiempo es inclemente, porque la lluvia o el frío hacen que sea más atractivo quedarse en casa. Si los datos convalidan esta hipótesis, como lo indica la línea C en la Figura 1.1., podría afirmarse que se ha explicado (al menos en parte) la varianza en la concurrencia a las urnas. Sin embargo, con ello no se habría propuesto *ninguna* explicación de por qué la línea C interseca el eje vertical en P y no en Q o en R. Es como si uno tomara el primer decimal como ya dado y se concentrara en explicar el segundo. A los efectos predictivos, con eso quizá sea suficiente. Pero a los efectos explicativos, es insuficiente. El «suceso en bruto» de que el 45% o más del electorado concurre habitualmente a votar *es* interesante, y clama por una explicación.

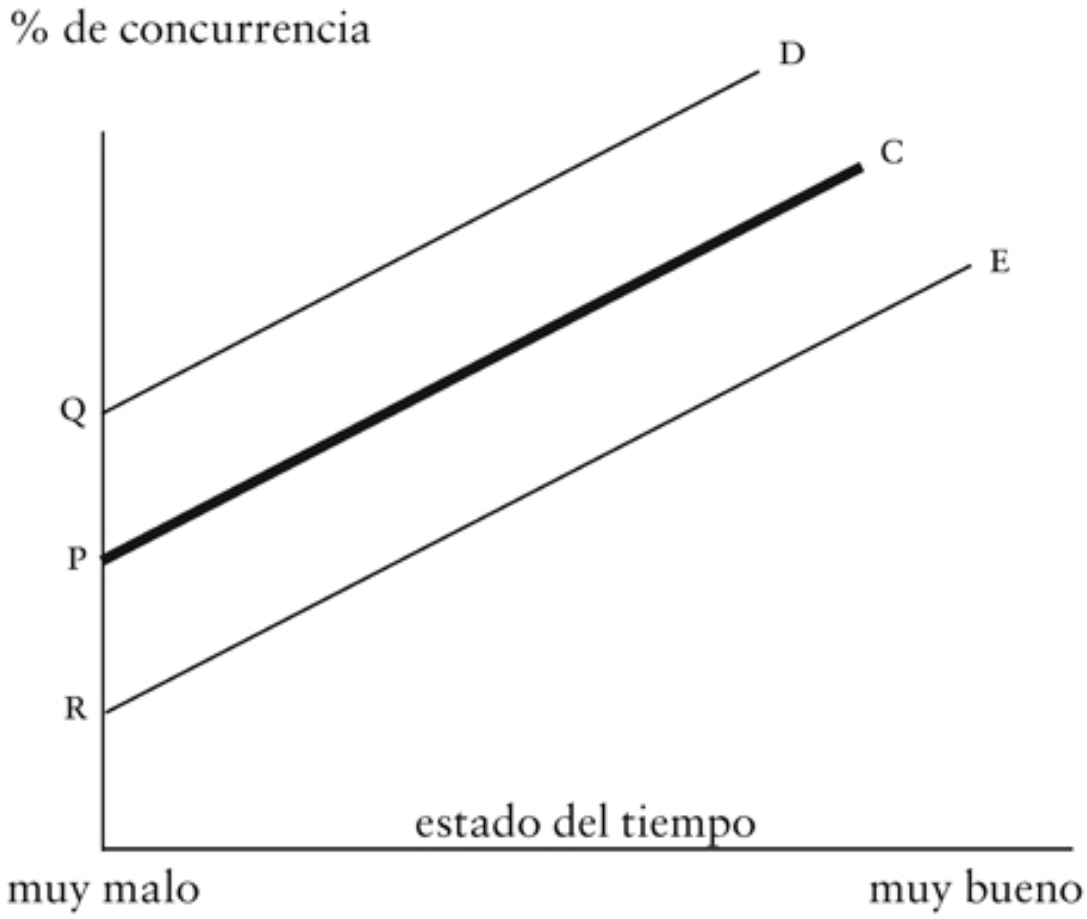


FIGURA 1.1

El procedimiento ideal, en una perspectiva de suceso-suceso, sería el siguiente. Considérense dos elecciones, A y B. Para cada una de ellas, identifíquense los sucesos que causan que un porcentaje dado de los votantes concurra a las urnas. Una vez que hemos explicado de tal modo la concurrencia a la elección A y la concurrencia a la elección B, la explicación de la diferencia (si la hay) se deduce automáticamente, como un subproducto. Como un beneficio adicional, tal vez podríamos explicar también si concurrencias idénticas en A y B son accidentales, esto es, debidas a diferencias que se compensan exactamente una a otra, o no. En la práctica, este procedimiento podría ser demasiado exigente. Los datos o las teorías disponibles quizá no nos permitirían explicar los fenómenos «en y por sí mismos». Deberíamos saber, sin embargo, que si recurrimos a

explicaciones de la varianza, nos embarcamos en una práctica explicativa que no es la mejor opción.

A veces, los especialistas en ciencias sociales tratan de explicar *no sucesos*. ¿Por qué mucha gente omite reclamar beneficios sociales si tiene derecho a ellos? ¿Por qué nadie llamó a la policía en el caso de Kitty Genovese?[6] Si consideramos la primera pregunta, la explicación podría ser que los individuos en cuestión *deciden* no reclamar sus beneficios debido al temor a ser estigmatizados o a preocupaciones por su autoestima. Como la toma de una decisión *es* un suceso, esa podría ser una explicación plenamente satisfactoria. Si fallara, los especialistas en ciencias sociales observarían una vez más las *diferencias* entre quienes tienen derecho a los beneficios y los reclaman y quienes, a pesar de disfrutar de ese mismo derecho, no lo ejercen. Supongamos que la única diferencia es que los últimos no saben que lo tienen. Como explicación, ésta es útil pero insuficiente. Para ir más allá, tendríamos que explicar *por qué* algunos individuos con derechos desconocen que los tienen. La comprobación de que, por ser analfabetos, no pueden leer las cartas que les informan de sus derechos también sería útil, pero insuficiente. En algún punto de la regresión explicativa, debemos o bien llegar a un suceso positivo, por ejemplo una decisión consciente de no aprender a leer y escribir o una decisión consciente de los funcionarios de retener información, o bien acudir a quienes procuran obtener los beneficios a los que están autorizados. Una vez explicado el comportamiento de estos últimos, la explicación de por qué los otros omiten reclamar su beneficio surgirá como un subproducto.

Cuando consideramos el caso de Kitty Genovese, vemos que no hay variación de la conducta que deba explicarse, dado que *nadie* llamó a la policía. Las descripciones del caso señalan que varios de los observadores *decidieron* no llamar. Desde el punto de vista de las causas próximas, esto proporciona una explicación plenamente satisfactoria, aunque tal vez queramos conocer las razones de esa decisión. ¿No telefonearon porque temían «verse involucrados» o porque cada uno de los observadores supuso que algún otro llamaría a la policía («Demasiados pastores poco vigilan»)? Sin embargo, algunos de ellos, al parecer, ni siquiera pensaron en hacer ese llamado. Un hombre y su mujer observaron el episodio como si fuera un espectáculo, mientras que otro hombre dijo que estaba cansado y se fue a la cama. Para explicar por qué no tuvieron una reacción más vigorosa,

podríamos aludir a la superficialidad de sus sentimientos, pero eso también sería dar razón de un *explanandum* negativo por medio de un *explanans* negativo. Su comportamiento, una vez más, sólo puede explicarse como un subproducto o un residuo. Si contamos con una explicación satisfactoria de por qué algunos individuos pensaron en llamar a la policía, aun cuando en definitiva hayan decidido no hacerlo, tendremos la única explicación que probablemente encontremos de por qué algunos otros ni siquiera consideraron esa posibilidad.

En el resto del libro suavizaré con frecuencia este enfoque purista o rigorista de lo que debe tomarse como un *explanandum* pertinente y una explicación apropiada. La insistencia en las explicaciones centradas en sucesos se asemeja un poco al principio del individualismo metodológico, que es otra de las premisas de este volumen. En principio, en las ciencias sociales las explicaciones deberían referirse únicamente a los individuos y sus acciones. En la práctica, los especialistas suelen referirse a entidades supraindividuales como las familias, las empresas o las naciones, sea como un *atajo inocuo* o como un *enfoque no ideal* que se han visto obligados a adoptar por falta de datos o teorías bien pulidas. Estas dos justificaciones también son válidas para el uso de hechos como *explananda* o *explanantia*, para las explicaciones de la varianza y no de los fenómenos «en y por sí mismos» y para el análisis de *explananda* negativos (no sucesos o no hechos). El objeto de la discusión precedente no es inducir a los especialistas en ciencias sociales a adoptar criterios inútiles o imposibles, sino sostener que, en el nivel de los primeros principios, el enfoque basado en los sucesos es intrínsecamente superior. Si los estudiosos tienen presente ese hecho, quizá puedan, al menos en algunas ocasiones, dar con mejores y más fructíferas explicaciones.

Es posible que a veces queramos explicar un suceso (o, mejor, un patrón de sucesos) por sus consecuencias y no por sus causas. No me refiero a la explicación por las consecuencias *intencionales*, dado que las intenciones existen con anterioridad a las elecciones o acciones que explican. La idea es más bien que los sucesos pueden explicarse por sus consecuencias *reales*: de manera habitual, sus consecuencias *beneficiosas* para alguien o algo. Como una causa debe preceder a su efecto, esta idea quizá parezca incompatible con la explicación causal. No obstante, ésta también puede adoptar la forma de la explicación por las consecuencias, si desde estas últimas hay una vuelta atrás hasta sus causas. En un inicio, un niño tal vez

llore simplemente porque siente dolor, pero si el llanto también despierta la atención de sus padres, aquél quizás empiece a llorar antes de lo que lo haría en otras circunstancias. En el Capítulo 16 y en el Capítulo 17 sostengo que este tipo de explicación es un tanto marginal en el estudio del comportamiento humano. En la mayor parte del libro me ocuparé de la variedad simple de la explicación causal, en la cual el *explanans* –que puede incluir creencias e intenciones orientadas hacia el futuro– es previo a la aparición del *explanandum*.[7]

Además de la forma plenamente respetable de la explicación funcional que se apoya en mecanismos específicos de realimentación, hay formas menos prestigiosas que se limitan a señalar la producción de consecuencias que son beneficiosas en algún aspecto, y luego, sin más argumentos, suponen que éstas bastan para explicar el comportamiento que las causa. Cuando el *explanandum* es un *rasgo distintivo*, como una única acción o suceso, este tipo de explicación fracasa por motivos puramente metafísicos. Para tomar un ejemplo de la biología, no podemos explicar la aparición de una mutación neutra o nociva señalando que fue la condición necesaria de otra, ésta de carácter ventajoso. Cuando el *explanandum* es un *tipo*, como un patrón de comportamiento recurrente, puede ser o no válido. Sin embargo, mientras no lo respalde un mecanismo específico de realimentación, debemos tratarlo como si fuera inválido. Los antropólogos han sostenido, por ejemplo, que la conducta vengativa tiene diversos tipos de consecuencias beneficiosas, que van desde el control demográfico hasta la imposición descentralizada de normas. (En el Capítulo 22 se encontrarán muchos otros ejemplos.) Suponiendo que esos beneficios se produzcan efectivamente, podría ser, de todos modos, que existieran por accidente. Para mostrar que surgen de manera no accidental, es decir, que sostienen la conducta vengativa que los causa, es indispensable la demostración de un mecanismo de realimentación. Y aun cuando se proponga uno de esos mecanismos, la aparición inicial del *explanandum* debe tener su origen en alguna otra cosa.

La estructura de las explicaciones

Querría hacer ahora una descripción más detallada de la explicación en las ciencias sociales (y, hasta cierto punto, en un marco más general). El primer paso es fácil de ignorar: antes de tratar de explicar un hecho o un

suceso, tenemos que establecer que el hecho *es* un hecho o que el suceso se ha producido efectivamente. Según escribió Montaigne: «Veo de ordinario que los hombres, en los hechos que se les presentan, prefieren ocuparse de buscar la razón que de buscar la verdad. […] Pasan por encima de los hechos, mas examinan con cuidado las consecuencias. Suelen comenzar así: ¿Cómo ocurre esto? Mas, ¿ocurre?, habríamos de decir».

De tal modo, antes de tratar de explicar, digamos, por qué hay más suicidios en un país que en otro, debemos cerciorarnos de que este último no tiende, quizá por razones religiosas, a denunciar menos de los que realmente ocurren. Antes de procurar explicar por qué España tiene una tasa de desempleo más elevada que Francia, tenemos que asegurarnos de que las diferencias informadas no se deben a distintas definiciones del paro o a la presencia de una gran economía informal en el primero de estos dos países. Si queremos explicar por qué el desempleo juvenil es más alto en Francia que en el Reino Unido, será necesario decidir cuál es el *explanandum*: la tasa de desempleo entre los jóvenes que buscan activamente trabajo o la tasa entre la juventud en general, incluidos los estudiantes. Si comparamos el paro en Europa y en los Estados Unidos, tenemos que decidir si el *explanandum* corresponde a los desempleados en sentido literal, lo que incluye a la población carcelaria, o en sentido técnico, que sólo incluye a quienes buscan trabajo.[8] Antes de explicar por qué la venganza adopta la forma de «ojo por ojo» (mato a uno de los tuyos cada vez que tú matas a uno de los míos), debemos verificar que eso es realmente lo que observamos, y no, por ejemplo, «dos ojos por uno» (mato a dos de los tuyos cada vez que tú matas a uno de los míos). Gran parte de la ciencia, incluida la ciencia social, procura explicar cosas que todos conocemos, pero la ciencia también puede hacer un aporte si establece que algunas de las cosas que todos creemos conocer sencillamente no son así, y agrega, por decirlo de algún modo, un fragmento de conocimiento en reemplazo del que ha sido eliminado.[9]

Supongamos ahora que tenemos un *explanandum* bien establecido para el cual no hay una explicación bien establecida: un *enigma*. El enigma puede ser un hecho sorprendente o contraintuitivo, o simplemente una correlación no explicada. Un ejemplo de pequeña escala es el siguiente: «¿Por qué en las bibliotecas de Oxford se hurtan más libros de teología que de otros temas?» Otro ejemplo de la misma escala, que un poco más adelante examinaré con mayor detalle, es: «¿Por qué hoy son más

numerosos que hace veinte años los espectáculos de Broadway que suscitan ovaciones de pie del público?»

En un plano ideal, los enigmas explicativos deben abordarse en una secuencia de cinco pasos que se detalla a continuación. En la práctica, sin embargo, los pasos 1, 2 y 3 a menudo aparecen en otro orden. Podemos jugar con diferentes hipótesis hasta que una de ellas se muestre como la más prometedora, y luego buscar una teoría que la justifique. Si los pasos 4 y 5 se cumplen como corresponde, podemos aún tener un alto nivel de confianza en la hipótesis preferida. No obstante, por razones que mencionaré al final del próximo capítulo, es posible que los estudiosos quieran limitar su libertad de elección entre hipótesis.

1. Elija la teoría (un conjunto de proposiciones causales interrelacionadas) que parezca prometer la explicación más fructuosa.
2. Especifique una hipótesis que aplica la teoría al enigma, en el sentido de que el *explanandum* se sigue lógicamente de la hipótesis.
3. Identifique o imagine descripciones verosímiles que puedan proponer explicaciones alternativas, también en el sentido de que el *explanandum* se sigue lógicamente de cada una de ellas.
4. Refute cada una de estas explicaciones antagónicas señalando implicaciones verificables adicionales que en realidad *no* se observan.
5. Fortalezca la hipótesis propuesta mostrando que tiene implicaciones verificables adicionales, preferentemente de «nuevos hechos», que se observan en concreto.

Estos procedimientos definen lo que suele denominarse *método hipotético deductivo*. En un caso dado, podrían tomar la forma mostrada en la Figura 1.2. Lo ilustraré con el enigma de la frecuencia creciente de las ovaciones de pie en Broadway. El dato no se basa en observaciones sistemáticas o experimentos controlados, sino en mis impresiones informales confirmadas por artículos periodísticos. A los presentes efectos, sin embargo, el endeble estatus del *explanandum* no tiene importancia. Si en nuestros días hay efectivamente más ovaciones de pie en Broadway que hace veinte años, ¿cómo podríamos intentar explicarlo?

Consideraré una explicación en función del precio creciente de las entradas a esos espectáculos neoyorquinos. Un diario menciona un comentario de Arthur Miller: «Me imagino que el público siente que, tras

haber pagado setenta y cinco dólares para sentarse, es hora de ponerse de pie. No pretendo ser cínico, pero probablemente todo cambió al aumentar el precio». Cuando la gente tiene que pagar setenta y cinco dólares o más por una platea, muchos no pueden decirse a sí mismos que el espectáculo ha sido malo o mediocre y que han dilapidado su dinero. Para confirmarse que la han pasado bien, aplauden como locos.

En términos más formales, la explicación se busca en la hipótesis «cuando la gente ha dedicado mucho dinero o esfuerzo a la obtención de un bien, tiende (en igualdad de las demás condiciones) a valorarlo más que cuando el precio es menor».[10] Dada la premisa fáctica de los precios crecientes, esta proposición pasa la prueba mínima que toda hipótesis explicativa debe cumplir: si es cierta, podemos inferir el *explanandum*. Pero esta prueba es verdaderamente mínima, y muchas proposiciones podrían satisfacerla.[11] Para fortalecer nuestra creencia en esta explicación en particular, debemos mostrar que está respaldada desde abajo, desde arriba y lateralmente. (Figura 1.2)

Una explicación recibe respaldo *desde abajo* si podemos deducir y verificar hechos observables sobre la base de la hipótesis, más allá del hecho que ésta pretende explicar. La hipótesis debe tener «capacidad explicativa en exceso». En el caso de los espectáculos de Broadway, cabría esperar que hubiera menos ovaciones de pie en aquellos cuyos precios, por algún motivo, no se han incrementado.[12] Esperaríamos además menos ovaciones de pie si grandes cantidades de entradas a un espectáculo se vendieran a empresas y éstas las cedieran a sus empleados. (Esto se consideraría como un «nuevo hecho».) Aun cuando esas entradas fuesen costosas, los espectadores no las han costeado de su bolsillo y, por lo tanto, no necesitan decirse que lo que reciben es dinero bien invertido.

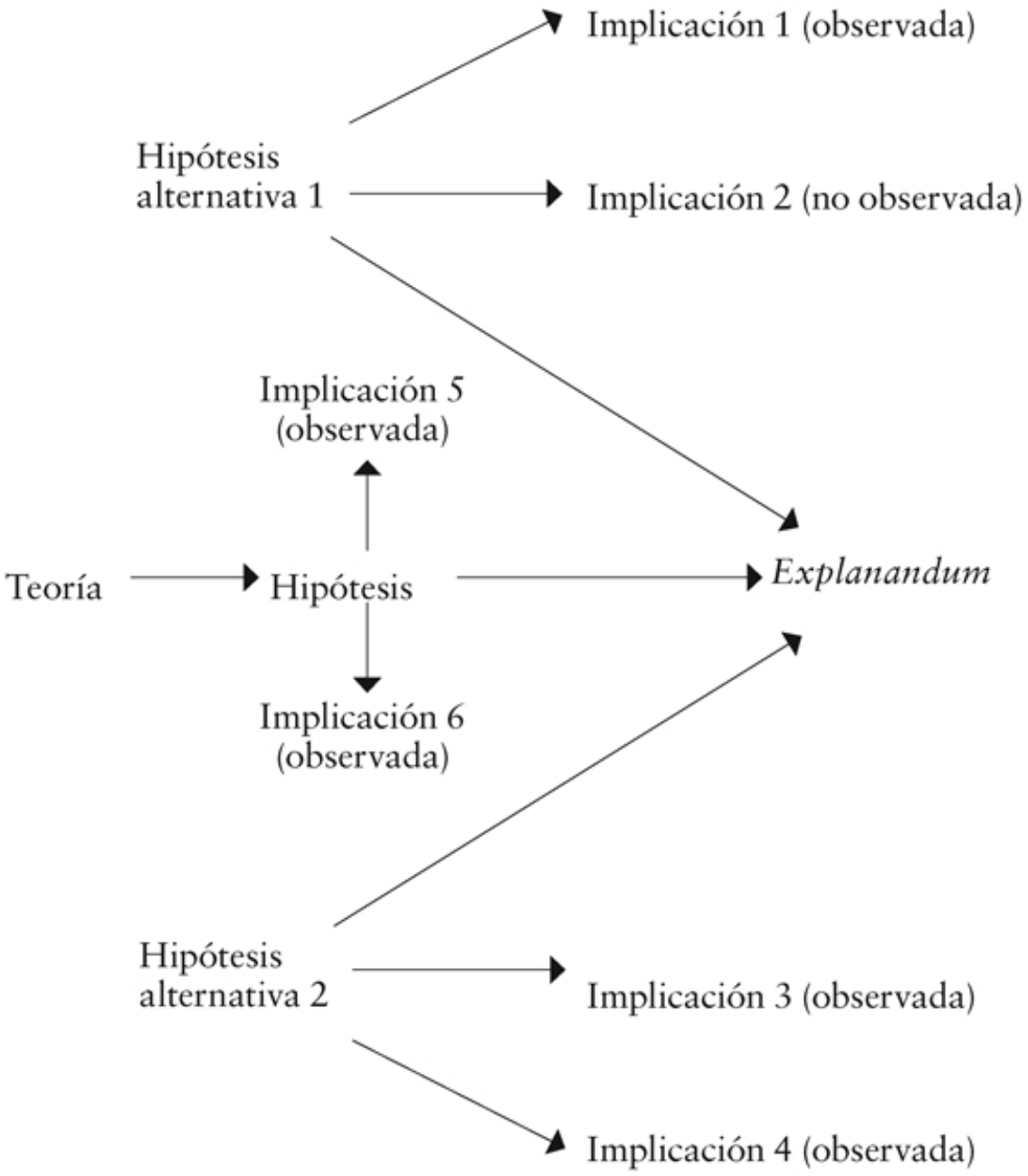


FIGURA 1.2

Una explicación recibe respaldo *desde arriba* si la hipótesis explicativa puede deducirse de una teoría más general.[13] En el presente caso, la proposición explicativa es una especificación de la teoría de la disonancia cognitiva propuesta por Leon Festinger. La teoría dice que cuando una

persona experimenta una inconsistencia o disonancia interna entre sus creencias y sus valores, podemos esperar algún tipo de reajuste mental que la elimine o la reduzca. Habitualmente, el ajuste tomará el camino de la menor resistencia. Una persona que ha gastado setenta y cinco dólares para ver un espectáculo que resulta ser malo no puede convencerse con facilidad de que ha pagado una suma menor. Le cuesta menos persuadirse de que, en realidad, el espectáculo es muy bueno.

Aunque no carece de problemas, la teoría de la disonancia cognitiva tiene fundamentos bastante sólidos. Su respaldo proviene en parte de casos que son muy diferentes de los que consideramos aquí, como sucede cuando una persona que acaba de comprar un automóvil busca con avidez avisos publicitarios de esa misma marca, para reforzar su convicción de que ha tomado una buena decisión. Otra parte del respaldo surge de casos muy similares, como cuando los dolorosos y humillantes rituales de iniciación de las fraternidades y sororidades universitarias inducen fuertes sentimientos de lealtad. No digo que la gente se diga de manera consciente: «Como he sufrido tanto para unirme a este grupo, la pertenencia a él debe ser algo bueno». El mecanismo mediante el cual el sufrimiento genera lealtad debe ser inconsciente.

Una explicación recibe *respaldo lateral* si podemos concebir y luego refutar explicaciones alternativas que también pasan la prueba mínima. Tal vez haya más ovaciones de pie porque el público de nuestros días, que llega en autobuses llenos desde Nueva Jersey, es menos sofisticado que el tradicional auditorio de hastiados residentes neoyorquinos. O quizá se deba a que los espectáculos son mejores que antes. Para cada una de estas alternativas, tenemos que concebir y después descartar hechos adicionales que existirían si aquellas fueran correctas. Si las ovaciones de pie son más frecuentes porque el público es más impresionable, también cabría esperar que lo hubieran sido en las actuaciones fuera de los circuitos céntricos veinte años atrás. Si los espectáculos son mejores que antes, habría que esperar que ese cambio se reflejara en las reseñas y el tiempo que permanecen en cartel.

En este procedimiento, el defensor de las hipótesis originales también tiene que ser el abogado del diablo. Uno debe *pensar coherentemente en contra de sí mismo*: hacerse las cosas lo más difíciles posible. Deberíamos elegir las explicaciones antagónicas alternativas más fuertes y verosímiles, en vez de buscar las que son de fácil refutación. Por razones similares,

cuando procuramos demostrar la capacidad explicativa excesiva de la hipótesis, debemos tratar de deducir y confirmar las implicaciones que son novedosas, contraintuitivas y tan diferentes del *explanandum* original como sea posible. Estos dos criterios –refutar las alternativas más verosí-miles y generar *nuevos hechos*– son decisivos para la credibilidad de una explicación. El respaldo desde arriba ayuda, pero nunca puede ser decisivo. A la larga, la teoría es respaldada por las explicaciones fructíferas que engendra, y no al revés. Emilio Segrè, un ganador del Premio Nobel de física, dijo que algunos laureados confieren honor al premio, mientras que otros lo obtienen de éste. Los últimos son, sin embargo, parasitarios de los primeros. De manera similar, una teoría es parasitaria de la cantidad de explicaciones fructíferas que genera. Si es capaz de otorgar respaldo a una explicación dada, sólo se debe a que lo ha recibido de explicaciones anteriores.

Lo que la explicación no es

Los enunciados cuyo propósito es explicar un suceso deben distinguirse de *otros siete tipos de enunciados*.

Primero, las explicaciones causales deben distinguirse de los *enunciados causales verdaderos*. Citar una causa no es suficiente: también es preciso señalar o al menos sugerir el mecanismo causal. En el lenguaje cotidiano, en las buenas novelas, en los buenos textos históricos y en muchos análisis de las ciencias sociales, el mecanismo no se menciona de manera explícita. Lo sugiere, en cambio, el modo de describir la causa. Cualquier suceso dado puede describirse de muchos modos. En las (buenas) explicaciones narrativas, se presupone tácitamente que, para identificar el suceso, sólo se utilizan los rasgos que tienen pertinencia causal. Si se nos cuenta que una persona ha muerto como consecuencia de haber comido alimentos en descomposición, suponemos que el mecanismo ha sido la intoxicación alimentaria. Si se nos dice que su muerte se ha producido como resultado de comer alimentos que le provocaban alergia, suponemos que el mecanismo ha sido una reacción alérgica. Supongamos ahora que la persona murió realmente debido a una intoxicación alimentaria, pero que también era alérgica a la comida en cuestión, la langosta. Decir que murió por haber comido un alimento al que era alérgica sería cierto, pero engañoso. Decir que murió por haber comido langosta sería cierto, pero

poco informativo. No sugeriría absolutamente ningún mecanismo causal y sería compatible con muchos: por ejemplo, que fue asesinada por alguien que había jurado matar al primer consumidor de langosta que viera.

Segundo, las explicaciones causales deben distinguirse de los enunciados referidos a las *correlaciones*. En ocasiones, estamos en condiciones de decir que un suceso de cierto tipo es invariable o habitualmente seguido por un suceso de otro tipo. Esto no nos permite decir que los sucesos del primer tipo causan sucesos del segundo, porque hay otra posibilidad: ambos podrían ser efectos comunes de un tercer suceso. En su *Life of Johnson*, James Boswell informa que un tal Macaulay, aunque «prejuicioso contra el prejuicio», afirmaba que cuando un barco llegaba a St. Kilda, en las Hébridas, «todos los habitantes cogen un resfriado». Si bien algunos propusieron una explicación causal de este (presunto) hecho, un corresponsal de Boswell informó a éste que «la situación de St. Kilda hace que, para que un extraño pueda atracar, sea indispensable un viento del nordeste. Es el viento, y no el extraño, el que ocasiona el resfriado epidémico». Considérese asimismo el descubrimiento de que los niños que son víctimas de casos contenciosos de custodia sufren más perturbaciones que los niños cuyos padres llegan a un acuerdo privado en ese tema. Podría ser que la disputa misma por la custodia explicase la diferencia, debido a que causa dolor y culpa en los niños. También podría suceder, no obstante, que los litigios por la custodia fueran más probables cuando los padres muestran una áspera hostilidad mutua, y que los hijos de padres de esas características tendiesen a ser perturbados. Para distinguir entre las dos interpretaciones, tendríamos que medir el sufrimiento antes y después del divorcio. Más adelante se contempla una tercera posibilidad.

A continuación, un ejemplo más complejo; a decir verdad, mi ejemplo favorito de este tipo de ambigüedad. En *La democracia en América*, Alexis de Tocqueville examina la presunta conexión causal entre el casamiento por amor y la infelicidad en el matrimonio. Señala que dicha conexión sólo existe en las sociedades en las que esos casamientos son la excepción, y los matrimonios arreglados son la regla. Únicamente la gente obstinada irá contra la corriente, y es poco probable que dos personas obstinadas disfruten de un matrimonio muy feliz.[14] Además, la gente que rema contra la corriente es maltratada por sus pares más conformistas, lo cual genera amargura e infelicidad. De estos argumentos, el primero descansa sobre una correlación no causal, debida a un «tercer factor», entre el casamiento por

amor y la infelicidad. El segundo apunta a una verdadera conexión causal, pero no la que los críticos de los matrimonios por amor a quienes Tocqueville dirigía su argumento tenían en mente. El casamiento por amor sólo provoca infelicidad en un contexto en que esta práctica es excepcional. Los biólogos suelen decir que esos efectos son «dependientes de la frecuencia».[15]

Por añadidura al problema del «tercer factor», la correlación puede suscitarnos incertidumbre con respecto a la *dirección* de la causalidad. Recordemos un viejo chiste:

> *Psicólogo*: Debería ser amable con Johnny. Proviene de un hogar destruido.
> *Maestro*: No me sorprende. No hay hogar que Johnny no sea capaz de destruir.

O, como ha dicho el comediante Sam Levinson: «La locura es hereditaria. Tus hijos pueden transmitírtela». Se da a entender así que un hijo perturbado puede llevar a los padres a divorciarse, y no que el divorcio causa la perturbación. De manera similar, una correlación negativa entre el grado de conocimiento que los padres tienen de las actividades de sus hijos adolescentes y la tendencia de éstos a meterse en líos no muestra por fuerza que la vigilancia parental funciona, sino únicamente que hay una escasa probabilidad de que los adolescentes propensos a meterse en líos mantengan a sus padres informados de sus actividades.

Tercero, las explicaciones causales deben distinguirse de los enunciados sobre la *necesitación*. Explicar un suceso es describir por qué ha ocurrido *tal y como* ha ocurrido. El hecho de que también podría haber ocurrido de alguna otra manera, y de que habría ocurrido de alguna otra manera de no haber ocurrido como ocurrió, no brinda una respuesta a la misma pregunta. Consideremos el caso de una persona que padece de cáncer de páncreas, enfermedad que ha de matarla sin lugar a dudas en un plazo de un año. Cuando el dolor se torna insoportable, la persona se suicida. Para *explicar* por qué ha muerto dentro de cierto período, carece de sentido decir que *tenía* que morir en ese plazo porque sufría de cáncer.[16] Si todo lo que sabemos sobre el caso consiste en la aparición del cáncer, la limitada esperanza de vida de las personas que padecen ese tipo de cáncer y la muerte de la persona en cuestión, es admisible inferir que ésta ha muerto a causa de la enfermedad. Tenemos el suceso anterior y un mecanismo causal suficiente para provocar el suceso posterior. Pero el mecanismo no es necesario: otro podría anticipársele. (En el ejemplo, la causa anticipante es

en sí misma un efecto de la causa anticipada, pero no es obligatorio que así sea; la persona también podría morir a raíz de un accidente automovilístico.) Para averiguar lo que sucedió realmente, necesitamos un conocimiento más sutil. La búsqueda nunca termina: hasta el último segundo, alguna otra causa podría anticiparse al cáncer.[17]

Los enunciados sobre la necesitación reciben a veces el nombre de «explicaciones estructurales». El análisis que hace Tocqueville de la Revolución francesa es un ejemplo. En su libro publicado sobre el tema, el autor menciona una serie de sucesos y tendencias desde el siglo XV hasta la década de 1780 y afirma que, en ese marco, la revolución era «inevitable». Con ello, probablemente quiere decir que 1) un número cualquiera de sucesos de pequeña o mediana magnitud habría bastado para desencadenarla, y 2) existía la virtual certeza de que ocurrirían *algunos* sucesos desencadenantes, aunque no necesariamente los que ocurrieron en concreto ni en el momento en que se produjeron. También parece argumentar que 3) luego de 1750 o tal vez de 1770, nadie podría haber hecho nada para impedir la revolución. Aunque Tocqueville dejó notas para un segundo volumen en el cual pretendía describir la revolución tal y como *había sucedido*, podríamos aducir que, si logró establecer con solidez los puntos 1, 2 y 3, ese paso adicional era innecesario. El inconveniente que presenta esta línea de razonamiento es que en muchas e interesantes cuestiones de las ciencias sociales (y en contraste con el ejemplo del cáncer), tesis como las enunciadas en esos tres puntos son muy difíciles de establecer con métodos que no estén teñidos por una visión retrospectiva.[18] Puede plantearse un argumento más convincente cuando sucesos similares ocurren independientemente unos de otros pero al mismo tiempo, lo cual sugiere que se los olía «en el aire». Un ejemplo de ello es el estudio de descubrimientos simultáneos en la ciencia.

Cuarto, la explicación causal debe distinguirse del *relato*. Una auténtica explicación describe lo sucedido, tal y como sucedió. Contar una historia es describir lo sucedido tal y como *podría* haber sucedido (y como tal vez sucedió). Acabo de sostener que las explicaciones científicas difieren de las exposiciones de lo que *tenía* que suceder. Digo ahora que también difieren de las descripciones de lo que *quizás* haya sucedido. La observación acaso parezca trivial o extraña. ¿Por qué querría alguien dar con una descripción puramente conjetural de un suceso? ¿Hay lugar en la ciencia para

especulaciones de este tipo? La respuesta es sí, pero su lugar no debe confundirse con el de la explicación.

El relato puede sugerir nuevas y parcas explicaciones. Supongamos que alguien afirma que el comportamiento abnegado o asistencial es prueba concluyente de que no todas las acciones son egoístas, y que el comportamiento emocional es prueba concluyente de que no todas las acciones son racionales. Uno podría llegar a la conclusión de que hay tres formas irreductiblemente diferentes de comportamiento: racional y egoísta, racional y no egoísta e irracional. La propensión a la parsimonia, que es característica de la buena ciencia, debería impulsarnos a cuestionar ese punto de vista. ¿No podría ser que, si la gente ayuda a otros, es porque espera reciprocidad, y si se encoleriza es porque esa reacción la ayuda a salirse con la suya? Al contar una historia acerca de la *probabilidad* de que el egoísmo racional genere un comportamiento altruista y emocional, podemos transformar un problema filosófico en un problema pasible de investigación empírica.[19] Una historia «así es porque así es» puede ser el primer paso en la construcción de una explicación convincente. De hecho, muchas de las «respuestas» que propongo en la conclusión a los enigmas que he presentado en la introducción tienen un fuerte aroma a ese tipo de historias.

Al mismo tiempo, los relatos pueden ser engañosos y nocivos si se los confunde con una explicación genuina. Con dos excepciones enunciadas en el siguiente párrafo, las explicaciones «como si» en realidad no explican nada. Considérese por ejemplo la afirmación corriente de que podemos utilizar el modelo de la elección racional para explicar el comportamiento, aun cuando *sepamos* que la gente no puede llevar a cabo los complejos cálculos mentales incorporados a dicho modelo (o en los anexos matemáticos a los artículos en los que éste se expone). Mientras el modelo proponga predicciones con un buen ajuste al comportamiento observado, tendremos derecho (se pretende) a suponer que los agentes actúan «como si» fueran racionales. Ésta es la concepción operacionalista o instrumentalista de la explicación, que tuvo su origen en la física y luego fue llevada por Milton Friedman a las ciencias sociales. La razón por la que podemos suponer que un buen billarista conoce las leyes de la física y puede efectuar mentalmente cálculos complejos es, se aduce, que ese supuesto nos permite predecir y explicar su comportamiento con gran exactitud. Preguntar por qué el supuesto es *verdadero* es no captar la idea.

Este argumento quizá sea valedero en algunas situaciones, en las cuales los agentes pueden aprender por ensayo y error a lo largo del tiempo. Sin embargo, es valedero justamente porque podemos señalar un *mecanismo* que produce de manera no deliberada el mismo resultado que un agente superracional podría haber calculado deliberadamente.[20] Enausencia de ese mecanismo, podríamos de todos modos aceptar la concepción instrumentalista si el supuesto nos permitiera predecir el comportamiento con una exactitud muy grande. La ley de la gravedad pareció misteriosa durante mucho tiempo, en cuanto se basaba aparentemente en la idea ininteligible de la acción a distancia. No obstante, gracias a hacer posibles predicciones que eran exactas hasta muchos decimales, la teoría de Newton fue aceptada sin discusiones hasta el surgimiento de la teoría de la relatividad general. El misterioso funcionamiento de la mecánica cuántica también se acepta, si bien no siempre sin desasosiego, porque permite hacer predicciones con una exactitud aún más increíble.

La ciencia social fundada en la elección racional no puede apoyarse en ninguno de estos dos respaldos. *No hay ningún mecanismo general no intencional que pueda simular o remedar la racionalidad*. El aprendizaje por refuerzo (Capítulo 16) puede hacerlo en algunos casos, aunque en otros genera desviaciones sistemáticas de la racionalidad. Algún tipo de análogo social a la selección natural podría hacerlo en otros casos, al menos de manera aproximada, de ser el índice de cambio del entorno menor que la velocidad de ajuste (Capítulo 17). En las situaciones únicas o en entornos rápidamente cambiantes, no conozco ningún mecanismo capaz de simular la racionalidad. Al mismo tiempo, el respaldo empírico a las explicaciones de fenómenos complejos basadas en la elección racional tiende a ser bastante débil. Ésta es desde luego una generalización excesiva. En vez de explicar qué quiero decir con «débil», me gustaría simplemente señalar el elevado nivel de discrepancia entre estudiosos competentes acerca de la capacidad explicativa de las hipótesis antagónicas. Aun en economía, que es en algunos aspectos la más desarrollada de las ciencias sociales, hay desacuerdos fundamentales y persistentes entre las «escuelas». *Jamás* observamos el tipo de precisión con muchos decimales que sea capaz de disipar la controversia.

Quinto, las explicaciones causales deben distinguirse de las *explicaciones estadísticas*. Si bien muchas explicaciones de las ciencias sociales tienen esta última forma, son insatisfactorias en cuanto no pueden

dar razón de sucesos individuales. La aplicación de generalizaciones estadísticas a los casos individuales es un grave error, no sólo en la ciencia sino en la vida cotidiana.[21] Supongamos que es cierto que los hombres suelen ser más agresivos que las mujeres. Decirle a un hombre airado que su ira es un producto de sus hormonas masculinas en vez de sostener que la ocasión no la justifica es cometer una falacia tanto intelectual como moral. La falacia intelectual consiste en presumir que una generalización válida para la mayoría de los casos lo es en cada caso. La falacia moral radica en tratar a un interlocutor como si lo gobernaran los mecanismos biológicos, en vez de considerar que está abierto al razonamiento y la argumentación.

Aunque las explicaciones estadísticas nunca son la primera opción, en la práctica es posible que no podamos apelar a nada mejor. Es importante señalar, con todo, que las guía inevitablemente el ideal de la mejor opción de la explicación causal. Parece ser un hecho estadístico que los ciudadanos de las democracias viven más que los ciudadanos de los regímenes no democráticos. Antes de llegar a la conclusión de que el régimen político explica la longevidad, quizá sería conveniente *controlar* otras variables que pueden ser responsables del resultado. Podría ser que hubiera más democracias que no democracias poseedoras de la propiedad X, y que en realidad fuera ésta la responsable de la esperanza de vida. Pero, puesto que la cantidad de esas propiedades es indefinida, ¿cómo sabremos qué variables controlar? La respuesta obvia es que necesitamos contar con la guía de una hipótesis causal. Parece verosímil, por ejemplo, que los ciudadanos de las sociedades industrializadas vivan más que los ciudadanos de sociedades menos desarrolladas. Si las sociedades industriales también tienden a ser más democráticas que los regímenes no industriales, el dato podría explicar los hechos observados. Para cerciorarnos de que el factor causal es la democracia y no la industrialización, tenemos que comparar regímenes democráticos y no democráticos igualmente industrializados, y ver si persiste una diferencia. Una vez razonablemente seguros de haber tomado en consideración otras causas posibles, también podemos tratar de averiguar *de qué manera* –a través de qué cadena o mecanismo causal– el tipo de régimen afecta la duración de la vida. Examino este segundo paso en el capítulo siguiente. Aquí, sólo quiero señalar que nuestra seguridad se basa inevitablemente en *intuiciones causales* acerca de los «terceros factores» verosímiles (y no verosímiles) que necesitamos tomar en consideración.[22]

Sexto, las explicaciones deben distinguirse de las *respuestas a «preguntas sobre el por qué»*. Supongamos que leemos un artículo académico y constatamos para nuestra sorpresa que el autor no hace mención de un texto importante y relevante, lo cual nos lleva a preguntarnos: «¿Por qué no lo cita?» Tal vez nuestra curiosidad quede perfectamente satisfecha si nos enteramos de que, en realidad, el autor desconocía esa obra anterior (aunque de ser así podríamos querer saber también por qué no ha hecho una exploración más exhaustiva de la literatura). Pero «no lo citó porque no lo conocía» no es una explicación. Leída como tal, la frase implicaría, absurdamente, la mención de un no suceso para explicar otro no suceso. Supongamos, no obstante, que comprobamos que el autor conocía el artículo pero *decidió* no citarlo porque él mismo no había sido mencionado en él. En ese caso, la respuesta al «por qué» también proporciona una explicación. Hay un suceso, la decisión de no citar el artículo, causado por un suceso anterior, la ira desencadenada por el hecho de no ser citado.

Para terminar, las explicaciones causales deben distinguirse de las *predicciones*. A veces podemos explicar sin ser capaces de predecir, y a veces predecimos sin ser capaces de explicar. Es cierto que en muchos casos una misma teoría nos permitirá hacer ambas cosas, pero creo que en las ciencias sociales esa situación es la excepción y no la regla.

Postergo hasta el próximo capítulo el principal análisis de por qué podemos tener capacidad explicativa sin contar con una fuerte capacidad predictiva. Como un breve anticipo, diré que la razón es que, en muchos casos, podemos identificar un mecanismo causal a posteriori, pero no predecir a priori cuál de varios mecanismos posibles se pondrá en marcha. El caso especial de la explicación biológica es un tanto diferente. Como lo discutiremos con mayor detalle en el Capítulo 16, la evolución es alimentada por los mecanismos gemelos de las mutaciones aleatorias y la selección (más o menos) determinista. Dado algún rasgo o patrón de conducta de un organismo, podemos explicar su *origen* recurriendo a un cambio aleatorio en el material genético y su *persistencia* debido a su efecto favorable sobre la aptitud reproductiva. No obstante, con anterioridad a la aparición de la mutación nadie la habría pronosticado. Por otra parte, como la aparición de una mutación restringe las mutaciones subsiguientes capaces de producirse, tal vez ni siquiera estemos en condiciones de predecir que una mutación determinada aparecerá tarde o temprano. Por eso las

explicaciones estructurales tienen escasas probabilidades de éxito en biología. El fenómeno de la *convergencia* (el desarrollo de adaptaciones similares en diferentes especies, porque éstas están sometidas a similares presiones ambientales) tiene un sabor estructural, pero no nos permite decir que las adaptaciones eran inevitables.

A la inversa, podemos tener capacidad predictiva sin capacidad explicativa. Para predecir que los consumidores comprarán en menor cantidad un bien cuando su precio aumente, no hace falta plantear una hipótesis con el fin de explicar su comportamiento. Sean cuales fueren las fuentes de la acción individual, racional, tradicional o simplemente fortuita, podemos predecir que, en general, la gente comprará menos ese bien por la sencilla razón de que les cuesta más solventarlo (Capítulo 9). En este caso hay varios mecanismos que llevarán forzosamente al mismo resultado, de modo que, con propósitos predictivos, no es necesario escoger entre ellos. No obstante, a los efectos explicativos, lo que cuenta es el mecanismo, pues nos permite entender, mientras que la predicción, a lo sumo, ofrece control.

Además, cuando se trata de predecir, la distinción entre correlación, necesitación y explicación resulta inútil. Si hay una regularidad legal entre un tipo de suceso y otro, no importa –con fines predictivos– que se deba a una relación causal entre ellos o al hecho de que sean efectos comunes de una tercera causa. En uno y otro caso, podemos utilizar la aparición del primer suceso para predecir la aparición del segundo. Nadie cree que los primeros síntomas de una enfermedad mortal causen la muerte, no obstante lo cual se los usa habitualmente para pronosticar ese suceso. De manera similar, si el conocimiento de la condición médica de una persona nos permite predecir que no estará viva al cabo de un año, el hecho de que el enfermo muera en un accidente automovilístico o se suicide porque la enfermedad es demasiado dolorosa no falsa la predicción.

* * *

NOTA BIBLIOGRÁFICA*

La concepción general de la explicación y la causación en la que me baso se expone con mayor detalle en Jon Elster, Dagfinn Føllesdal y Lars Walløe, *Rationale Argumentation: ein Grundkurs in Argumentationsund Wissenschaftstheorie*, Berlín, De Gruyter, 1988 (traducción inglesa en preparación). Para aplicaciones a la acción humana, remito al lector a

Donald Davidson, *Essays on Actions and Events*, Nueva York, Oxford University Press, 1980 [*Ensayos sobre acciones y sucesos*, Barcelona y México, Crítica/Instituto de Investigaciones Filosóficas de la UNAM, 1995]. Mi crítica de la explicación funcional se expone en varios lugares, sobre todo, en *Explaining Technical Change: A Case Study in the Philosophy of Science*, Cambridge y Nueva York, Cambridge University Press, 1983 [*El cambio tecnológico: investigaciones sobre la racionalidad y la transformación social*, Barcelona, Gedisa, 1992]. Se encontrarán detalles del caso de Kitty Genovese en Abraham M. Rosenthal, *38 Witnesses: The Kitty Genovese Case*, Berkeley, University of California Press, 1999. Un acceso conveniente a las concepciones de Leon Festinger es Stanley Schachter y Michael Gazzaniga (comps.), *Extending Psychological Frontiers: Selected Works of Leon Festinger*, Nueva York, Russell Sage, 1989. Los ejemplos de los efectos «de hijos a padres» proceden de dos estimulantes libros de Judith Rich Harris, *The Nurture Assumption: Why Children Turn Out the Way They Do*, Nueva York, Free Press, 1998 [*El mito de la educación: por qué los padres pueden influir muy poco en sus hijos*, Barcelona, Grijalbo, 1999] y *No Two Alike*, Nueva York, Norton, 2006. Examino los puntos de vista de Tocqueville sobre la causalidad en «Patterns of causal analysis in Tocqueville's *Democracy in America*», *Rationality and Society*, 3(3), 1991, págs. 277-297, y sus concepciones sobre la Revolución francesa en «Tocqueville on 1789: preconditions, precipitants, and triggers», en Cheryl Welch (comp.), *The Cambridge Companion to Tocqueville*, Cambridge y Nueva York, Cambridge University Press, 2006, págs. 49-80. La defensa de Milton Friedman de la racionalidad del «como si» en «The methodology of positive economy» (1953) se reeditó en May Brodbeck (comp.), *Readings in the Philosophy of the Social Sciences*, Londres, Macmillan, 1969, págs. 508-528 [«La metodología de la economía positiva», en Milton Friedman, *Ensayos sobre economía positiva*, Madrid, Gredos, 1967]. Una reciente defensa del enfoque del «como si» en las ciencias políticas es Rebecca B. Morton, *Methods and Models: A Guide to the Empirical Analysis of Formal Models in Political Science*, Cambridge y Nueva York, Cambridge University Press, 1999. Como la mayoría de los partidarios del enfoque, la autora no dice por qué razón debemos creer en la ficción del «como si». Se encontrará una excepción parcial en Debra Satz y John Ferejohn, «Rational choice and social theory», *Journal of Philosophy*, 91(2), 1994, págs. 71-87. El examen de las «preguntas sobre el por qué» se

basa en Bengt Hansson, «Why explanations? Fundamental, and less fundamental ways of understanding the world», *Theoria*, 72(1), 2006, págs. 23-59. La independencia de la ley de la demanda con respecto a los supuestos motivacionales se ha señalado en Gary Becker, «Irrational behavior in economic theory», *Journal of Political Economy*, 70(12), 1962, págs. 1-13.

Notas al pie

[1] Para anticiparnos a una distinción analizada más adelante, señalemos que Carter no *omitió [*fail*] intentar*, sino que *intentó y fracasó* [*failed*]. Una no acción como la omisión de un intento no puede tener eficacia causal, excepto en el sentido indirecto de que, si los otros perciben o infieren que el agente omite actuar, tal vez emprendan acciones que de lo contrario no emprenderían.

[2] Otra de sus ilustraciones es la concurrencia de los votantes a las urnas, que se analiza más adelante.

[3] Las respuestas fluctúan. Además, la cantidad de gente que apoya la pena de muerte para los asesinatos disminuye de manera drástica cuando se menciona como alternativa la prisión perpetua sin posibilidad de libertad bajo palabra.

[4] Como dicen a veces los economistas, sólo les interesa lo que pasa «en el margen».

[5] Estrictamente hablando, la cadena causal podría encaminarse en el otro sentido, de las actitudes a la conducta, pero en este caso la hipótesis no es verosímil.

[6] El 27 de marzo de 1964, en Queens, Nueva York, treinta y ocho ciudadanos respetables y respetuosos de la ley observaron durante más de media hora cómo un asesino merodeaba y apuñalaba a una mujer en tres ataques separados ocurridos en Kew Gardens. En dos ocasiones, sus conversaciones y el súbito resplandor de las luces de sus dormitorios interrumpieron al hombre y lo hicieron marcharse atemorizado. El atacante regresó dos veces, detectó a la mujer y volvió a apuñalarla. Ni una sola persona telefoneó a la policía durante los ataques; un testigo llamó luego de que la mujer muriera.

[7] A algunos efectos, acaso sea útil distinguir entre explicaciones causales, intencionales y funcionales. La física sólo se vale de la explicación causal; la biología también admite la explicación funcional; y las ciencias sociales agregan a las dos anteriores la explicación intencional. Sin embargo, en el nivel más fundamental todas las explicaciones son causales.

[8] En cualquiera de estos dos últimos casos, algunos individuos tal vez decidan ser delincuentes o estudiantes porque no creen que, de intentarlo, sean capaces de obtener un empleo. A algunos efectos, querríamos a lo mejor contarlos entre los parados; a otros efectos, los excluiríamos.

[9] Así como puede contribuir a explicar las creencias populares en no hechos, la ciencia puede ayudar a explicar las creencias populares en las falsas explicaciones. Por ejemplo, la mayoría de los pacientes de artritis creen que el mal tiempo desencadena el dolor artrítico. Sin embargo, los estudios realizados indican que esa conexión no existe. Tal vez deberíamos dejar a un lado la búsqueda del vínculo causal entre el mal tiempo y el dolor artrítico y tratar, en cambio, de explicar por qué los artríticos creen que lo hay. Con toda probabilidad, alguna vez se les dijo que había una conexión y después prestaron mayor atención a los casos que confirmaban la creencia que a aquellos que la desmentían.

[10] A veces se utiliza una idea similar para defender los elevados honorarios de los psicoterapeutas: los pacientes no creerán en la terapia a menos que paguen mucho por ella. Pero, que yo sepa, ningún terapeuta ha declarado que dona el 50% de sus honorarios a la Cruz Roja.

[11] La mente humana parece tener una tendencia a convertir esta exigencia mínima en una exigencia suficiente. Una vez que damos con una explicación que *puede* ser cierta, no solemos tomarnos el tiempo de someterla a pruebas adicionales o de considerar explicaciones alternativas. La elección de una explicación tal vez se deba a la idea del *post hoc ergo propter hoc* (después de esto, por lo tanto, a consecuencia de esto) o a una inferencia que, a partir del hecho de que una explicación determinada es *más verosímil que otras*, llega a la conclusión de que tiene *mayores probabilidades* de ser correcta.

[12] *No* deberíamos esperar necesariamente que menos espectadores se pusieran de pie en los sectores más baratos. Es posible que esos espectadores consideraran una tontería permanecer

sentados mientras los demás se levantan; además, quizá tuvieran que pararse para ver a los actores, cuya visión, de lo contrario, quedaría bloqueada por quienes están de pie delante de ellos.

[13] Para decirlo con mayor exactitud: si es una *especificación* de una teoría más general. La relación entre una teoría general y una hipótesis explicativa específica rara vez es de carácter deductivo. Por un lado, quizás haya algún defecto en la teoría misma (véase el Capítulo 2); por otro, una teoría dada puede, por lo común, operacionalizarse de muchas maneras.

[14] Aquí, el «tercer factor» es un rasgo de carácter, la obstinación, y no un suceso.

[15] El primer mecanismo es un *efecto de selección*, y el segundo, un auténtico *efecto secundario*. La validez de la distinción es bastante amplia. Si nos preguntamos por qué alguien puesto en una situación determinada (por ejemplo, dedicarse a cierta ocupación, estar desempleado o estar internado a causa de una enfermedad mental) tiene más probabilidades de permanecer en esa misma situación cuanto más tiempo haya pasado en ella, podría ser que uno de los dos mecanismos (o ambos) estuvieran en acción. Los desempleados de larga data, por ejemplo, podrían formar un subconjunto de la población dotado de calificaciones para las que hay escasa demanda; como alternativa, todos los individuos empleados podrían tener iguales probabilidades de perder su trabajo, pero una vez que lo han perdido, el hecho de estar desempleados los modifica (o modifica la percepción que de ellos tienen los empleadores), de modo que su probabilidad de reingresar al mercado laboral disminuye con el transcurso del tiempo. La «teoría del etiquetamiento» de la enfermedad mental o el delito se apoya en el (dudoso) supuesto de que los efectos secundarios predominan sobre los efectos de selección.

[16] James Fitzjames Stephen escribe que «la ley determina con absoluta claridad que, si en razón de [una] agresión, [un hombre] ha muerto en la primavera por una enfermedad que debería haberlo llevado a la tumba en, digamos, el verano, dicha agresión es la causa de su muerte».

[17] Es preciso distinguir la anticipación causal de la sobredeterminación causal. Una ilustración de esta última es una persona alcanzada al mismo tiempo por dos balas, cada una de las cuales habría sido suficiente para matarla. La primera se ilustra con el caso de una persona muerta con una sola bala, que se anticipa a la acción de otra disparada unos segundos después.

[18] La revolución norteamericana es quizás un candidato más adecuado para una explicación estructural. Un agudo observador neutral como el ministro francés Choiseul advirtió ya en 1765 que la independencia de las colonias norteamericanas era inevitable. Para un comentarista francés imparcial como Raymond Aron, la independencia de Argelia también era una conclusión predeterminada mucho antes de producirse. La Revolución francesa es más afín al derrumbe del comunismo, ineluctable sobre todo en retrospectiva.

[19] En este caso en particular, las historias «así es porque así es» [*just-so stories*] resultan falsas, dado que las personas también ayudan a otros en interacciones únicas, y la ira puede hacer que otros se abstengan de interactuar con ellas.

[20] En esta primera aparición en el libro de la palabra «agente», quizá merezca la pena señalar que muchos estudiosos prefieren hablar de «actor». Tal vez los economistas piensen en términos de agentes, y los sociólogos, en términos de actores. Aunque en realidad no importa qué término utilicemos, prefiero «agente» porque sugiere agencia; «actor», en contraste, sugiere una audiencia que puede o no estar presente.

[21] La falacia inversa (el uso de un caso individual para producir o respaldar una generalización) también debe evitarse. Proust escribió que Françoise, ama de llaves de la familia del narrador, «se inclinaba tanto a tomar lo particular por lo general como lo general por lo particular». Esta combinación puede ser perniciosa. Supongamos que observamos a un miembro del grupo X decir una mentira. Generalizamos y nos forjamos la creencia de que los miembros de dicho grupo tienden a mentir. A continuación, al observar a otro integrante del grupo, suponemos que está mintiendo. Por último, el supuesto (no verificado) se utiliza como prueba adicional de la generalización.

[22] Por ejemplo, no hay un mecanismo causal verosímil que nos impulse a tomar como variable de control el tamaño de la población de los regímenes democráticos y no democráticos. Aunque no se puede excluir un vínculo causal entre dicho tamaño y la duración promedio de la vida, las ciencias sociales no han verificado ninguna conexión de esas características; tampoco puedo imaginar una que no sea artificiosa.

* Las versiones en castellano, entre corchetes, de los títulos citados por el autor han sido buscadas e incluidas en éste y en todos los demás capítulos por el traductor de esta obra. [T.]

Capítulo 2

Mecanismos

Abriendo la caja negra

Los filósofos de la ciencia suelen argumentar que una explicación debe basarse en una *ley general*. Explicar un suceso es referirse a un conjunto de condiciones generales junto con un enunciado en el sentido de que, cada vez que existen estas condiciones, se sigue un suceso de ese tipo. En este capítulo planteo dos objeciones a la idea, una moderada y relativamente poco polémica, la otra más radical y expuesta a la discusión.

La primera objeción es que aun cuando podamos establecer una ley general para deducir de ella el *explanandum* (la segunda objeción niega que podamos hacerlo en todas las circunstancias), esto no siempre equivale a una explicación. Podemos remitir una vez más a la distinción entre la explicación, por un lado, y la correlación y la necesitación, por otro. Una ley general que sostenga que ciertos síntomas de una enfermedad son ineluctablemente seguidos por la muerte no explica por qué ha muerto la persona. Una ley general basada en la naturaleza fundamental de la enfermedad no explica la muerte, si un suicidio o un accidente automovilístico se han anticipado a aquélla.

Para burlar estos problemas, a menudo se sostiene que debemos reemplazar la idea de ley general por la de *mecanismo*. En realidad, como más adelante utilizo el término «mecanismo» en un sentido especial, utilizaré la expresión «cadena causal» para denotar lo que evoco aquí.[23] En vez de tratar de explicar un suceso E mediante el enunciado «cada vez que

ocurren los sucesos C1, C2, …, Cn, se sigue un suceso de tipo E», quizá podamos procurar establecer la cadena causal que lleva de las causas C1, C2, …, Cn a E. Con frecuencia se alude a este paso como la «apertura de la caja negra». Supongamos que sabemos que los grandes fumadores tienen muchas más probabilidades que otros de padecer cáncer de pulmón. Este hecho podría deberse o bien a que el hábito de fumar es una causa de cáncer pulmonar, o bien a que la gente predispuesta a fumar también tiene predisposición al cáncer (tal vez los genes que predisponen al cáncer de pulmón están vinculados con los genes que llevan a alguna gente a adquirir más rápidamente la adicción a la nicotina).[24] Para establecer la primera explicación, tendremos que mostrar una cadena de relaciones fisiológicas de causa-efecto que comience con el consumo abundante de tabaco y termine con el cáncer de pulmón. La explicación final será más sutil, tendrá más vínculos causales y será más convincente que el enunciado de caja negra «el hábito de fumar provoca cáncer».

Otra posibilidad: supongamos que alguien ha afirmado que el desempleo elevado provoca guerras de agresión y ha aducido pruebas en favor de una conexión con carácter de ley entre los dos fenómenos. Insistamos: ¿cómo podemos saber que se trata de un efecto causal y no de una mera correlación? ¿No podría decirse que los altos índices de fertilidad, que generan desempleo, también motivan a los líderes políticos a desatar guerras agresivas? Las guerras infructuosas disminuirían al menos la cantidad de población, y las victoriosas redundarían en nuevos territorios para la expansión y la migración. Para eliminar esta posibilidad, en primer lugar tendríamos que tomar como variable de control el índice de fertilidad (y otros «terceros factores» verosímiles) y ver si la conexión persiste. Aunque lo hiciera, no quedaríamos todavía satisfechos hasta que se nos permitiera echar un vistazo a la caja negra y se nos dijera *de qué manera* el desempleo elevado causa guerras. ¿La causa radica en que el paro induce a los líderes políticos a buscar nuevos mercados por medio de la guerra? ¿O en que creen que aquél genera una agitación social que debe canalizarse hacia un enemigo externo, para impedir los movimientos revolucionarios en el propio suelo? ¿O en que están convencidos de que la industria armamentística puede absorber el desempleo? ¿O será acaso que los desempleados tienden a votar a líderes populistas propensos a eludir la diplomacia y apelar en cambio a las guerras para resolver los conflictos?

Consideremos con mayor detalle la última alternativa. ¿*Por qué* votarían los desempleados por líderes populistas irresponsables y no por políticos pertenecientes a uno de los partidos establecidos? Una vez más, es posible imaginar unas cuantas maneras de abrir esta caja negra en particular. Tal vez la clientela de los políticos populistas se incline más a votar cuando está desempleada, porque el coste de oportunidad de su concurrencia a las urnas (es decir el valor de su tiempo) es menor que cuando tienen un trabajo. O quizá los líderes populistas muestren mayor tendencia a proponer soluciones instantáneas al problema del paro. O bien propicien políticas que castigan a aquellos a quienes los desempleados creen responsables o beneficiarios de su penosa situación, se trate de capitalistas o de minorías étnicas económicamente exitosas.

Examinemos ahora con más detenimiento esta última propuesta. ¿*Por qué* querrían los desempleados castigar a los capitalistas o a las minorías adineradas? ¿No es éste un enunciado más de caja negra? Una manera de comprenderlo consistiría en afirmar que el móvil de los desempleados es el egoísmo material. Si el Estado pudiese confiscar la riqueza de esas elites, los fondos podrían utilizarse en una redistribución que los beneficiara. O tal vez estén motivados por un deseo de venganza, que los incita a castigar a la elite aun cuando con ello no se beneficien en términos materiales. Si se estima que los ricos promueven implacables reducciones de personal para incrementar sus ganancias, quienes pierden su trabajo pueden usar las urnas para desquitarse. O bien los desempleados podrían estar simplemente envidiosos de los miembros de la minoría inteligente que tienen éxito donde ellos han fracasado, por lo cual utilizan las urnas para bajarles los humos.

Que yo sepa, el desempleo elevado no causa guerras de agresión. Todo el ejercicio es hipotético. Creo, no obstante, que respalda la idea de que la credibilidad de una explicación aumenta en la medida en que las leyes generales se enuncian en términos de una cadena causal. En el nivel de esas leyes generales, nunca podemos tener la certeza de que hemos tomado en consideración todos los «terceros factores» relevantes. Siempre puede haber alguna causa al acecho entre bastidores que dé razón tanto del *explanandum* como de su presunta causa. Si aumentamos el número de eslabones de la cadena causal, reducimos ese peligro.

Sin embargo, no es posible eliminarlo. La especificación de una cadena causal no significa el abandono completo de las leyes generales; sólo implica ir de leyes generales en un alto nivel de abstracción a leyes en un

nivel más bajo. Podríamos, por ejemplo, reemplazar la ley universal «el desempleo elevado causa guerras» por las leyes menos abstractas «los líderes populistas son propensos a la guerra» y «los desempleados votan a líderes populistas». Esta última ley, a su vez, podría ser reemplazada por la conjunción de «los desempleados envidian a las minorías ricas» y «quienes envidian a las minorías ricas votan a líderes populistas». Como en el caso de cualquier otra ley, éstas tal vez resulten ser meras correlaciones. Si la envidia a las minorías y el desempleo son efectos comunes de una causa conjunta, el éxito electoral de los líderes propensos a la guerra se debería no al paro sino a un factor causalmente correlacionado con él. Sin embargo, en este nivel más pormenorizado hay menos factores para controlar. Cuanto más nos concentremos en la historia causal, más fácil será cerciorarnos de que no estamos ante una mera correlación.

Las explicaciones en términos de leyes (muy) generales también son insatisfactorias, por ser demasiado opacas. Aun cuando se presenten con un argumento blindado sobre la existencia de un vínculo universal entre el desempleo y las guerras de agresión y la afirmación convincente de que se han controlado todos los «terceros factores» remotamente verosímiles, podríamos querer todavía saber *de qué manera* el desempleo causa guerras. Tal vez creamos que la explicación es correcta y, sin embargo, no estemos satisfechos con ella. Como he indicado en el capítulo anterior, ése era el estatus de las explicaciones basadas en la ley de la gravedad antes de la relatividad general. La acción a distancia era tan misteriosa que muchos se negaban a creer que pudiera ser la última palabra. Como la ley permitía corregir predicciones con una precisión de muchos decimales, los escépticos tuvieron que aceptar que las cosas pasaban «como si» aquélla fuera cierta, aunque no admitieran la existencia de una fuerza que podía «actuar donde no estaba».

Mecanismos

Los lectores bien pueden haber pensado que los ejemplos de presuntas leyes universales expuestos en este ejercicio son bastante inverosímiles. Estoy de acuerdo. En parte, su falta de verosimilitud puede deberse a los límites de mi imaginación a la hora de urdirlos, pero creo que también hay razones más profundas. En las ciencias sociales hay, simplemente, muy pocas leyes generales bien establecidas. La «ley de la demanda» (cuando

los precios suben, los consumidores compran menos) tiene una base sólida, pero en cuanto ley es bastante débil.[25] La ley de la gravedad, por ejemplo, no sólo nos dice que cuando la distancia entre dos objetos se incrementa la fuerza de atracción entre ellos disminuye: nos dice *cuánto* disminuye (inversamente con el cuadrado de la distancia). En las ciencias sociales no hay nada parecido a la ley de la gravedad.[26]

La ley de la demanda y la ley de Engel, según la cual la proporción del ingreso destinado a la compra de alimentos disminuye cuando el ingreso aumenta, son lo que podríamos llamar *leyes débiles*. Para cada cambio (hacia arriba o hacia abajo) en la variable independiente, nos permiten predecir la *dirección* o el signo de un cambio (hacia arriba o hacia abajo) en la variable dependiente. No nos permiten, sin embargo, predecir la *magnitud* del cambio. Aunque débiles, esas leyes tienen algún contenido, dado que gracias a ellas podemos excluir toda una gama de valores posibles de la variable dependiente. No nos ayudan, no obstante, a identificar el valor que se realizará dentro de la gama no excluida.

La ley de la demanda no sólo es débil; también es muy inadecuada a los fines explicativos. Como vimos en el Capítulo 1, es compatible con varios supuestos sobre el comportamiento de los consumidores. Para *explicar* por qué estos últimos reducen la compra de un bien cuando éste se encarece, tendríamos que adoptar y someter a prueba un supuesto específico sobre las reacciones individuales de los consumidores a los cambios de precio. La palabra clave es «individual». En las ciencias sociales, una explicación satisfactoria debe basarse en última instancia en hipótesis sobre el comportamiento individual. Este principio, conocido como «individualismo metodológico», es la premisa de todo este libro. Da a entender que la psicología y tal vez la biología deben tener una importancia fundamental en la explicación de los fenómenos sociales. Si vacilo con respecto a la biología, no es porque la crea en principio incapaz de explicar aspectos del comportamiento humano, sino porque me parece que está demasiado poco desarrollada para cumplir esa tarea.

Para explicar el comportamiento individual, tenemos que apoyarnos sobre todo en lo que llamo *mecanismos*. A grandes rasgos, los mecanismos son *patrones causales de aparición frecuente y fáciles de reconocer que se ponen en funcionamiento en condiciones generalmente desconocidas o con consecuencias indeterminadas*. Nos permiten explicar, pero no predecir. Se ha sostenido, por ejemplo, que por cada niño que cae en el alcoholismo en

respuesta a un entorno alcohólico, otro elude el alcohol en respuesta a ese mismo entorno. Ambas reacciones encarnan mecanismos: hacer lo que hacen los padres y hacer lo contrario de lo que ellos hacen. No podemos decir con antelación qué será del hijo de un alcohólico, pero si termina por ser o bien abstemio o bien alcohólico, quizá sospechemos que sabemos por qué.

No afirmo que actúe aquí ningún tipo de indeterminación objetiva; en rigor de verdad, ese concepto tiene poco sentido fuera de la mecánica cuántica. Sólo pretendo decir que a menudo podemos explicar el comportamiento demostrando que es una instancia de un patrón causal general, aunque no seamos capaces de explicar por qué ha aparecido éste. Los mecanismos del conformismo (por ejemplo, hacer lo que hacen nuestros padres) y del anticonformismo (hacer lo contrario de lo que ellos hacen) son muy generales. Si podemos mostrar que el comportamiento de un hijo de padre alcohólico es un ejemplo de uno u otro mecanismo, habremos proporcionado una explicación del comportamiento. Cabría objetar que, mientras no demostremos por qué el hijo se convirtió en alcohólico (digamos) y no en abstemio, no habremos explicado nada. Yo coincidiría, sin duda, en que una explicación que demostrara por qué se produjo un resultado y no otro sería mejor, y no niego que en ocasiones podríamos darla. Pero subsumir un ejemplo individual en un patrón causal más general también es suministrar una explicación. Saber que el hijo se convirtió en alcohólico como consecuencia del conformismo es disipar parte de la opacidad del resultado, aunque algo de ésta se mantendrá mientras no expliquemos también por qué aquél adoptó una conducta conformista.

Dije que un mecanismo es «un patrón causal de aparición frecuente y fácil de reconocer». La sabiduría popular proverbial ha identificado muchos de esos patrones.[27] Según mi definición predilecta: «Un proverbio se ha transmitido a lo largo de muchas generaciones. Resume, en una frase breve, un principio general o una situación común, y cuando lo decimos, todo el mundo sabe exactamente a qué nos referimos». Por otra parte, los proverbios suelen enunciar mecanismos (en el sentido utilizado aquí) y no leyes generales. Considérese, en particular, su llamativa tendencia a aparecer en pares mutuamente excluyentes. Por un lado tenemos «la ausencia aviva el amor», pero por otro, «ojos que no ven, corazón que no siente». Por un lado tal vez creamos que el fruto prohibido sabe mejor, pero

por otro, que están verdes las uvas que no podemos alcanzar. Por un lado, «Dios los cría y ellos se juntan», pero por otro, «los opuestos se atraen». Por un lado, «de tal palo tal astilla», pero por otro, «de padre mezquino, hijo pródigo». Por un lado, «vísteme despacio que estoy apurado», pero por otro, «vacilar es perderse». Por un lado, «recordar una desdicha es reavivarla», pero por otro, «el recuerdo de peligros pasados es agradable». (Como se señala más adelante, los dos últimos no son, en realidad, mutuamente excluyentes.) Podríamos mencionar muchos otros ejemplos.

Hay muchos pares de mecanismos opuestos que los proverbios no parecen abarcar. Considérese, por ejemplo, lo que podemos llamar par derrame-compensación. Si una persona muy trabajadora se va de vacaciones, ¿habrá que esperar que traslade el mismo ritmo frenético a sus actividades de esparcimiento (efecto de derrame) o, al contrario, que se relaje por completo (efecto de compensación)? ¿O deberemos presuponer que los ciudadanos de las democracias son propensos o adversos a la religión? Si trasladan el hábito de decidir por sí mismos de la esfera política a la esfera religiosa (derrame), cabría esperar creencias religiosas débiles. Si la falta de una autoridad superior en política los lleva a buscar la autoridad en otra parte (compensación), un régimen político democrático tendería antes bien a favorecer la religión. Un interrogante contemporáneo, que aún parece no resuelto, es si la violencia en televisión estimula la violencia en la vida real (derrame) o la atenúa (compensación).

Mecanismos similares pueden aplicarse a las relaciones entre individuos. Considérese la cuestión de explicar las donaciones a las instituciones de beneficencia. Un individuo tal vez se preocupe sobre todo por la eficiencia del donar. Si otros dan poco, su donación tendrá mayor impacto y, por lo tanto, él se sentirá más inclinado a dar; si otros dan mucho, su donación importará menos y quizá se abstenga de hacerla. Otro donante puede interesarse más en la equidad (entre donantes). Si los demás dan poco, no verá razones para dar más; a la inversa, si los otros dan mucho, tal vez se sienta obligado a seguir su ejemplo. El mismo par de mecanismos puede actuar en situaciones de acción colectiva. Cuando un movimiento popular crece, es probable que algunos individuos lo abandonen, porque ya no creen que su actividad signifique mucho, en tanto que otros tal vez se incorporen a él, pues les parece impropio mantenerse a un costado mientras otros pagan el coste (Capítulo 24).

Incluso los proverbios que no se emparejan con otro de sentido opuesto suelen expresar mecanismos y no leyes. El proverbio «aun los mejores nadadores se ahogan» sería absurdo si consideráramos que significa que la propensión a ahogarse aumenta de manera invariable con la aptitud natatoria. Sin embargo, en el caso de algunos nadadores bien puede suceder que la confianza en sus aptitudes para la natación aumente más rápidamente que éstas y los lleve a correr riesgos injustificados («el orgullo antecede a la caída»). Considérese asimismo un proverbio al que volveré varias veces en este libro: «Somos muy crédulos con lo que esperamos y con lo que tememos».[28] Aunque el proverbio es poco creíble si se lo toma al pie de la letra, como una ley universal, es útil para recordarnos que, además del conocido fenómeno del pensamiento desiderativo, hay una propensión menos conocida a lo que podríamos llamar *pensamiento contramotivado*.[29] Consideremos, para terminar, las afirmaciones proverbiales «demasiados pastores poco vigilan» y «demasiados cocineros salan mucho la sopa».* Una vez más, el valor del proverbio no consiste en enunciar una ley universal, sino en sugerir mecanismos. El primer proverbio podría ser cierto si cada pastor creyera que todos los demás hacen guardia (recuérdese el caso Kitty Genovese), y el segundo, si cada cocinero creyese que ninguno de los otros ha salado la sopa.

Al definir los mecanismos, también dije que «se ponen en funcionamiento en condiciones generalmente desconocidas o con consecuencias indeterminadas». La mayoría de los mecanismos proverbiales que he mencionado hasta aquí se incluyen en la primera categoría. No sabemos qué condiciones desencadenarán el conformismo o el anticonformismo, el pensamiento desiderativo o el pensamiento contradesiderativo (contramotivado), las preferencias adaptativas (las uvas verdes) o las preferencias contraadaptativas (el prado es más verde). Sabemos que a lo sumo se realizará un miembro de cada par, pero no podemos decir cuál. El modificador «a lo sumo» es importante, porque algunas personas pueden no estar sujetas a ninguno de los miembros de estos pares de mecanismos. La autonomía genuina implica no ser conformista ni anticonformista. Las creencias de la gente pueden ser independientes de sus deseos, y éstos, independientes de sus oportunidades.

En otros casos, los proverbios indican el desencadenamiento simultáneo de dos mecanismos con efectos en direcciones opuestas sobre el resultado. Cuando así sucede, la indeterminación radica en determinar el *efecto neto*

de los mecanismos, y no en determinar cuál de ellos (si lo hay) se pondrá en funcionamiento. Considérese, por ejemplo, «la necesidad es la madre del ingenio» y «cuesta caro ser pobre». El primer proverbio afirma un vínculo causal entre la pobreza y un fuerte *deseo* de innovación, y el segundo, un vínculo entre la pobreza y la escasez de *oportunidades* de innovación. Como el comportamiento está conformado tanto por deseos como por oportunidades (Capítulo 9), en general no podemos decir si el impacto neto de la pobreza sobre la innovación es positivo o negativo. Considérese asimismo el par de proverbios mencionados antes, «recordar una desdicha es reavivarla» contra «el recuerdo de peligros pasados es agradable». El primero de ellos se basa en lo que se ha dado en llamar «efecto de dotación»: el recuerdo de una mala experiencia es una mala experiencia.[30] El segundo descansa en un «efecto de contraste»: el recuerdo de una mala experiencia realza el valor del presente.[31] En general, no podemos decir si el efecto neto de una mala experiencia anterior sobre el bienestar posterior será positivo o negativo.

Insistamos: no nos limitamos a los proverbios. Tómense en consideración, por ejemplo, dos mecanismos no proverbiales que intervienen en lo que se ha denominado «psicología de la tiranía». Si el tirano incrementa la opresión de los súbditos, es probable que se produzcan dos efectos. Por un lado, castigos más duros los disuadirán de resistir o rebelarse. Por otro, cuanto más se comporte el tirano como tal, más lo odiarán los súbditos. Como cualquier pendenciero, es probable que inspire a la vez miedo y odio. Si éste predomina sobre aquél, la opresión terminará por volverse en su contra. En los países ocupados por los alemanes durante la Segunda Guerra Mundial, los miembros de la resistencia a veces explotaban este mecanismo cuando mataban soldados enemigos para provocar una represalia, con la idea de que el «efecto de tiranía» prevalecería sobre el efecto de disuasión.[32] Véase asimismo el caso un tanto similar de una persona a quien una barrera o un impedimento separan de su meta. Esta amenaza a su libertad de acción puede inducir lo que los psicólogos llaman «reactancia», una motivación para recuperar o restablecer esa libertad. Los efectos de la barrera y la consecuente reactancia se oponen entre sí, y en general no podemos decir cuál será más fuerte.[33] Como ilustración, piénsese en el efecto que provoca en un varón pequeño el hecho de que sus padres escondan un tambor porque no quieren que lo toque.[34]

Aun cuando conozcamos el efecto neto, tal vez no estemos en condiciones de explicarlo. Supongamos que, de algún modo, hemos logrado observar y medir un efecto neto cero de los efectos de dotación y contraste en lo referido a una buena experiencia del pasado. Este resultado podría alcanzarse de dos maneras. Aunque la comida francesa de tres estrellas que comí el año pasado redujo el placer ulteriormente obtenido con otras experiencias en restaurantes franceses más corrientes, este efecto negativo sobre mi bienestar tiene una compensación exacta en el recuerdo de lo fantástica que fue esa comida. No obstante, la observación de un efecto neto igual a cero también es perfectamente compatible con los efectos cero de dotación y contraste, así como con el caso en que ambos son muy fuertes y de igual magnitud. Mientras no sepamos cuál es el caso en cuestión, no podremos afirmar haber explicado el resultado. Para evaluar la fuerza de cada efecto, podríamos observar el resultado en una situación en que no se espera que el otro se produzca. Si, como parece verosímil, el placer que me causa la cocina griega no es afectado por la comida francesa de tres estrellas, podemos identificar la fuerza del mero efecto de dotación.

Una indeterminación conexa puede surgir con respecto al primer tipo de mecanismos, los desencadenados en «condiciones generalmente desconocidas». Volvamos a examinar el caso de un padre alcohólico. Tomemos la población total de alcohólicos (o una muestra representativa amplia) y supongamos que, en promedio, sus hijos no beben ni más ni menos que los hijos de no alcohólicos. Si, para simplificar, pasamos por alto la influencia de los factores genéticos, este descubrimiento hipotético podría entenderse de dos maneras. Por un lado, podría ser que los hijos de alcohólicos no fueran ni conformistas ni anticonformistas: esto es, su conducta en materia de bebida podría estar conformada por las mismas causas que la conducta de los hijos de no alcohólicos. Por otro lado, podría ser que la mitad de los hijos de alcohólicos fueran conformistas y la otra mitad, anticonformistas, con lo cual el efecto neto sería igual a cero.

De manera similar, las teorías sobre el comportamiento electoral han identificado tanto un mecanismo de causa perdida como un mecanismo de causa triunfante. Quienes están sometidos al primero tienden a votar por el candidato que está detrás en las encuestas preelectorales, mientras que los sujetos al segundo votan por el favorito. Si ambos tipos se mezclan en forma pareja, tal vez no haya un efecto neto perceptible y, de ese modo, las encuestas serán buenos predictores del voto real. Sin embargo, la falta de

influencia de las encuestas sobre el conjunto de la votación no demostraría que los individuos son indiferentes a ellas. Los débiles efectos agregados de la violencia televisiva sobre la violencia en la vida real podrían enmascarar fuertes efectos opuestos sobre determinados subgrupos. En todos estos casos, un agregado neutral podría reflejar o bien una población homogénea de individuos no afectados o una población heterogénea de individuos fuertemente afectados, pero en direcciones opuestas. La necesidad de disipar esta ambigüedad proporciona un argumento más a favor del individualismo metodológico. Para explicar el comportamiento en el nivel agregado, debemos observar el comportamiento de los componentes individuales.

Mecanismos moleculares

He considerado hasta aquí lo que podríamos llamar mecanismos «atómicos»: reacciones psicológicas elementales que no pueden reducirse a otros mecanismos del mismo nivel. Bien podríamos preguntarnos hasta dónde nos llevarán estos mecanismos psicológicos en la explicación de los fenómenos sociales. La respuesta es que podemos utilizar los mecanismos atómicos como unidades de construcción de mecanismos «moleculares» más complejos. Tal vez convenga comenzar una vez más con proverbios. Dos de ellos rezan: «el miedo es a menudo más grande que el peligro» y «cuanto más miedo, más peligro». Tomados en conjunto, dan a entender que el miedo excesivo puede ser su propia justificación. Un proverbio inglés dice que «en todo rebaño hay una oveja negra». Un proverbio francés nos cuenta que «hace falta una sola oveja descarriada para arruinar un rebaño». Si los juntamos, podemos inferir que todos los rebaños están arruinados.[35]

Considérese por ejemplo un cuerpo colegiado autónomo, como un departamento universitario o una cooperativa de trabajadores. El siguiente escenario es lo bastante frecuente e inteligible para hacer las veces de mecanismo molecular. Primero, por las leyes de la probabilidad, cualquier grupo de veinte o más miembros puede contener al menos una persona gárrula y díscola, un «arruinador», definido por el diccionario como «aquel que echa a perder la oportunidad de victoria de un adversario, sin ser él mismo un potencial ganador». Segundo, en un grupo que contiene una persona de estas características, el autogobierno colectivo es muy difícil.

Las discusiones son eternas; hay un cuestionamiento constante de las decisiones previas; un espíritu de formalismo reemplaza la colegialidad informal; se genera mal humor, etc. A la larga, el grupo recibirá con agrado la propuesta de que el gobierno quede a cargo de un comité ejecutivo más pequeño e incluso de un solo individuo.

Dejemos atrás los proverbios y examinemos otro mecanismo molecular. Durante siglos o milenios, las elites fueron circunspectas con respecto a la democracia como forma de régimen, porque creían que daría pábulo a todo tipo de comportamientos peligrosos y licenciosos. Sin embargo, las oportunidades de comportamiento peligroso no lo producen por sí solas: también debe haber un motivo. ¿Podrían los regímenes democráticos restringir de algún modo los deseos de los ciudadanos de hacer lo que la democracia les permite hacer? Ésa era la idea de Tocqueville, quien creía que, para satisfacer una necesidad de autoridad que la política no saciaba, los ciudadanos democráticos acudirían a la religión, la cual tiende a limitar y restringir lo que éstos desean. Los críticos de la democracia se equivocaban, sostenía, porque se concentraban únicamente en las oportunidades y hacían caso omiso de los deseos. Si bien expuso este argumento como si resultara una ley universal, es más plausible entenderlo en términos de mecanismos. Por una parte, si actúa el efecto de derrame y no el de compensación, la falta de autoridad política debilitará la religión en vez de fortalecerla. Por otra, aun cuando esté en acción el primero de esos efectos, no podemos llegar a ninguna conclusión acerca del efecto neto. Si el conjunto de oportunidades se expande mucho y los deseos se restringen sólo un poco, el efecto neto de la democracia quizá consista en aumentar y no reducir la incidencia del comportamiento en cuestión.

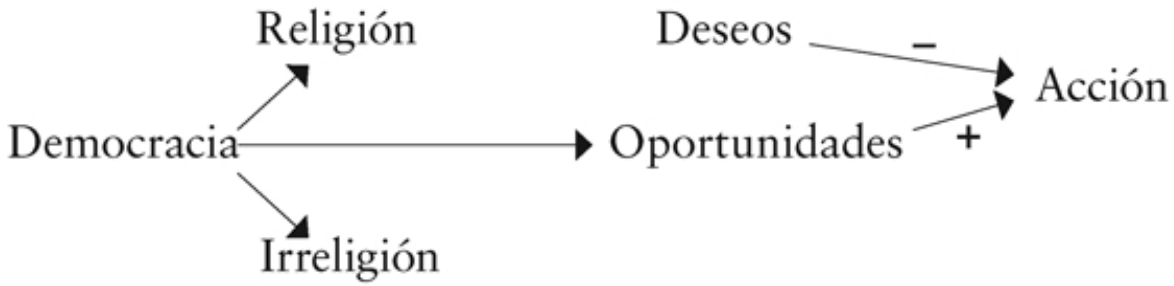


FIGURA 2.1

Los dos pares de mecanismos se representan en forma sumaria en la Figura 2.1. *Si* la influencia de la democracia sobre la religión es mediada por el efecto de compensación y no por el efecto de derrame, las sociedades democráticas serán religiosas. *Si* el efecto negativo de la democracia sobre los deseos (mediado por la religión) es lo bastante fuerte para compensar su efecto positivo sobre las oportunidades, los ciudadanos democráticos exhibirán un comportamiento moderado.[36]

Mecanismos y leyes

Con frecuencia, la explicación con referencia a mecanismos es lo mejor que podemos hacer, pero en ocasiones hay cosas más adecuadas. Una vez que identificamos un mecanismo que «se pone en funcionamiento en condiciones generalmente desconocidas», quizá seamos capaces de identificar las condiciones desencadenantes. En ese caso, el mecanismo será reemplazado por una ley, si bien de ordinario ésta será débil, en el sentido antes definido.

El sentido común supone que un regalo despertará el agradecimiento de su receptor. Si no es así, lo censuramos. Los moralistas clásicos –de Montaigne a La Bruyère– argumentaban en cambio que los regalos suelen suscitar el resentimiento y no el agradecimiento de sus receptores. Al parecer, tanto el sentido común como los moralistas están bien encauzados, pero no nos dicen cuándo podemos esperar uno u otro resultado. Un moralista de la Antigüedad clásica, Publio Siro, enunció *condiciones desencadenantes*: un pequeño obsequio crea una obligación, uno grande crea un enemigo.[37] Al recurrir al tamaño del regalo como condición desencadenante, hemos transformado el par de mecanismos en algo parecido al enunciado de una ley.[38] Para citar otro ejemplo, tal vez podríamos enunciar cuándo se resuelve una tensión entre un deseo y una creencia («disonancia cognitiva») al modificar esta última y cuándo al modificar el primero.[39] Las creencias puramente fácticas pueden ser demasiado recalcitrantes para ser de fácil modificación (Capítulo 7). La persona que ha pagado setenta y cinco dólares por una entrada a un espectáculo de Broadway no puede engañarse con facilidad y creer que ha pagado cuarenta dólares. Sin embargo, habitualmente podrá encontrar algunos aspectos atractivos del espectáculo y convencerse de que son más importantes que sus aspectos fallidos.

Antes me referí al contraste entre el mecanismo del «fruto prohibido» y el mecanismo de las «uvas verdes». En algunos casos, quizá podamos predecir cuál se pondrá en funcionamiento. En un experimento, se pidió a sujetos de una condición que clasificaran cuatro discos en función de su atractivo y se les dijo que al día siguiente recibirían uno de ellos, elegido al azar. Sujetos de otra condición también ordenaron los discos según el mismo criterio y se les indicó que al día siguiente podrían elegir uno de ellos. Llegado ese día, se informó a todos los participantes que el disco que habían clasificado en tercer lugar no se podía conseguir, y se les pidió que volvieran a clasificar los cuatro, como parte de un intento de descubrir cómo podía la segunda escucha de un disco afectar su evaluación. Tal y como lo pronostica la teoría de la reactancia, los sujetos de la primera condición exhibieron el efecto de las «uvas verdes» al reducir el valor de la opción inasequible, mientras que los sujetos de la segunda condición mostraron el efecto del «fruto prohibido», al elevarlo. (Un grupo de control al que nada se dijo sobre la eliminación no mostró cambios.) La diferencia crucial es que el segundo grupo experimentó una amenaza a su libertad, cosa que no sucedía con el primero.

Querría examinar, sin embargo, un ejemplo más complejo. Con respecto al par de proverbios «la ausencia aviva el amor» y «ojos que no ven, corazón que no siente», hay en realidad un tercer proverbio que sugiere una condición desencadenante: «una breve ausencia puede hacer mucho bien». La Rochefoucauld propuso una condición diferente: «La ausencia mengua las pasiones moderadas e intensifica las grandes, así como el viento apaga una vela pero aviva un incendio». Estas proposiciones plausibles no son leyes muy fuertes. Para poder predecir el curso de la pasión, tendríamos que saber a qué se llama breve ausencia (¿tres semanas?) y pasión intensa (¿la que no nos permite pegar un ojo durante la noche?). Además, sería preciso especificar cómo *interactúan* la duración de la ausencia y la intensidad de la pasión para generar el aumento o la disminución de esta última durante una ausencia. Me gustaría seguir con esta cuestión.

Interacción entre causas

En general, las ciencias sociales no son muy buenas para explicar de qué manera interactúan las causas con el fin de producir un efecto conjunto. Lo más común es suponer que cada causa contribuye por separado al efecto

(un «modelo aditivo»). Para explicar el ingreso, por ejemplo, podemos suponer que se debe en parte al ingreso de los padres y en parte a su educación, y luego utilizar métodos estadísticos para determinar los aportes relativos de estas dos causas. Este enfoque podría no ser adecuado para el ejemplo que he examinado. La duración de la ausencia tal vez no haga una contribución separada a la intensidad de la emoción posterior a ella; antes bien, su efecto podría depender de la intensidad de la emoción previa a la ausencia. Este efecto de interacción se muestra en la Figura 2.2

Algunos estudiosos sostienen, no obstante, que el mundo o, al menos, la parte de éste que ellos estudian, sencillamente no muestra muchas interacciones de este tipo. Es muy poco habitual, afirman, que para los niveles bajos de la variable independiente X, la variable dependiente Z aumente (disminuya) con la variable dependiente Y, mientras que para los niveles altos de X un incremento en Y cause una disminución (aumento) en Z. La relación hipotetizada en la Figura 2.2 sería (de existir) una excepción. A lo sumo, argumentan estos estudiosos, lo que comprobamos es que en los niveles bajos de X, Y tiene escaso efecto sobre Z, en tanto que sí tiene un efecto en niveles más elevados de X. Al explicar el ingreso, por ejemplo, uno puede suponer que el ingreso parental contribuye más o menos en diferentes niveles de educación de los padres. Este tipo de interacción puede mostrarse mediante un término de interacción multiplicativa, de modo que Z sea una función de X, Y y XY. En contraste, la *inversión* del efecto causal de Y sobre Z en niveles más altos de X no puede mostrarse de esta manera. Sin embargo, si podemos dar crédito a la idea de que esas inversiones son raras, no hará falta preocuparse demasiado por ellas.

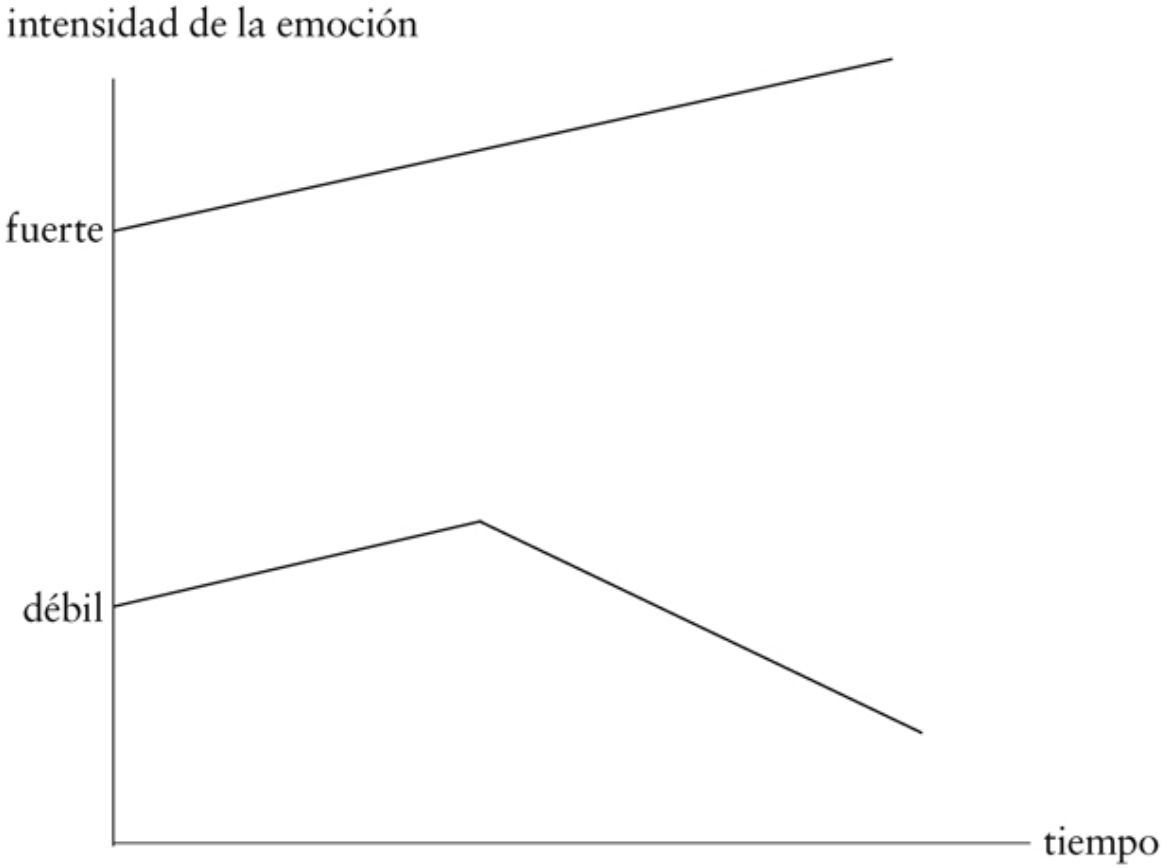


FIGURA 2.2

La existencia de un efecto de interacción puede estar sujeta al mismo tipo de indeterminación que encontramos más generalmente en los mecanismos. Tómese en consideración la interacción entre la edad y las actitudes políticas básicas como causas de extremismo. Cabría suponer que las organizaciones juveniles serán la izquierda de los partidos mismos, lo cual daría a la Juventud Conservadora un matiz más leve de azul. Como alternativa, las organizaciones juveniles de los partidos políticos serán más extremas que los propios partidos: la Juventud Conservadora tendrá un matiz más oscuro de azul y la Juventud Socialista un matiz más intenso de rosa. Ambas conjeturas parecen verosímiles y ambos patrones, de hecho, se han observado. Considérese asimismo la interacción entre el humor previo al consumo y la ingesta de drogas como causas del humor posterior a ésta. Uno podría suponer que drogas como el alcohol y la cocaína *mejoran el humor*, al atenuar la depresión y convertir el contento en euforia. Pero

también sería lícito sospechar que las drogas *multiplican los humores*, al empeorar los malos y mejorar los buenos. Una vez más, ambas conjeturas parecen plausibles y ambos patrones se atestiguan. En los dos casos, el primer mecanismo es compatible con un modelo aditivo, mientras que el segundo implica un efecto de inversión.

Cuando nos enfrentamos con datos recalcitrantes, el agregado de un término de interacción, o «ajuste de curvas», no es la única respuesta posible. Hay una estrategia alternativa, la de la «explotación de datos». En un ejercicio de ajuste de curvas, mantenemos fijas las variables dependiente e independiente y salimos de compras, por así decir, en busca de una función matemática que proporcione un buen ajuste estadístico. En un ejercicio de explotación de datos, mantenemos fija la función matemática (por lo común un modelo aditivo simple) y recorremos tiendas en busca de variables independientes que tengan un buen ajuste con la variable dependiente. Supongamos que por «buen ajuste» nos referimos a una correlación que tenga sólo un 5% de probabilidades de ocurrir por azar. En cualquier estudio de un fenómeno social complejo como el ingreso, es fácil enumerar una docena de variables con posibilidades verosímiles de afectarlo.[40] Además, hay probablemente media docena de maneras diferentes de conceptualizar el ingreso. Sería muy improbable que ninguna de las variables independientes mostrara en el nivel del 5% una correlación con una de sus definiciones.[41] Las leyes de la probabilidad nos dicen que la más improbable coincidencia sería que las improbables coincidencias nunca sucedieran.[42]

Una vez que el investigador ha identificado una función matemática conveniente o un conjunto de variables dependientes o independientes adecuadas, puede comenzar a buscar una historia causal para proponer una intuición como respaldo de los descubrimientos. Cuando pone por escrito los resultados para su publicación, con frecuencia la secuencia se invierte. El investigador manifestará que empezó con una teoría causal; luego buscó la forma más plausible de transformarla en una hipótesis formal, y por último comprobó que los datos la confirmaban.[43] Esto es ciencia espuria. En las ciencias naturales no es necesario que la «lógica de la justificación» coincida con la «lógica del descubrimiento» o la refleje. Una vez planteada una hipótesis en su forma definitiva, su génesis es irrelevante. Importa lo que tiene por delante, sus consecuencias, no lo que ha quedado atrás, sus orígenes. Es así porque la hipótesis puede someterse a prueba en una

cantidad indefinida de observaciones, más allá de las que en un principio inspiraron al investigador a pensar en ella. En las ciencias sociales (y las humanidades), la mayoría de las explicaciones se valen de un conjunto finito de datos. Como los procedimientos de recolección de éstos no suelen estar estandarizados, es posible que los investigadores se vean ante la imposibilidad de verificar sus hipótesis contra nuevos datos.[44] Y si los procedimientos *están* estandarizados, los datos pueden no reflejar una realidad cambiante. Es imposible explicar los patrones de consumo, por ejemplo, sin tomar en cuenta la aparición de nuevos productos y las modificaciones de precios de los anteriores.

No hay duda de que hay prácticas bien definidas de este tipo. No sé cuán habituales son; sólo sé que tienen la suficiente difusión para generar preocupación entre los especialistas cuidadosos en ciencias sociales. La principal razón del problema es quizá nuestra inadecuada comprensión de la causalidad multifactorial. Si tuviéramos vigorosas intuiciones sobre la interacción de varias causas para producir un efecto, no sería necesario apoyarse en el procedimiento mecánico de «agregar un término de interacción» cuando un modelo aditivo fracasa. No obstante, como nuestras intuiciones son débiles, no sabemos realmente qué buscar y, en consecuencia, la manipulación de los modelos parece ser la única alternativa, al menos si persistimos en la ambiciosa meta de proporcionar explicaciones con carácter de ley. Vistos los peligros de la manipulación, tal vez lo mejor sería limitar nuestras ambiciones.

* * *

NOTA BIBLIOGRÁFICA

Muchas de las ideas de este capítulo son una adaptación de lo expuesto en el Capítulo 1 de mi *Alchemies of the Mind: Rationality and the Emotions*, Cambridge y Nueva York, Cambridge University Press, 1999 [*Alquimias de la mente: la racionalidad y las emociones*, Barcelona, Paidós, 2002]. En ese capítulo también menciono obras de Raymond Boudon, Nancy Cartwright y Paul Veyne que hacen propuestas similares. Una exposición reciente es la de Peter Hedström, *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge y Nueva York, Cambridge University Press, 2005. También se encontrarán reflexiones provechosas sobre los mecanismos psicológicos en Fritz Heider, *The*

Psychology of Interpersonal Relations, Hillsdale (Nueva Jersey), Lawrence Erlbaum, 1958, y Robert P. Abelson, *Statistics as Principled Argument*, Hillsdale (Nueva Jersey), Lawrence Erlbaum, 1995 [*La estadística razonada: reglas y principios*, Barcelona, Paidós, 1998]. Este último también hace cuerdas e ingeniosas observaciones sobre las trampas del análisis estadístico. Una breve exposición clásica de la idea de que la ciencia explica mediante leyes generales es Carl Hempel, *Philosophy of Natural Science*, Englewood Cliffs (Nueva Jersey), Prentice-Hall, 1966 [*Filosofía de la ciencia natural*, Madrid, Alianza, 1973]. El principio del individualismo metodológico se aborda de manera exhaustiva en la cuarta parte de May Brodbeck (comp.), *Readings in the Philosophy of the Social Sciences*, Londres, Macmillan, 1969, y en la sexta parte de Michael Martin y Lee C. McIntyre (comps.), *Readings in the Philosophy of Social Science*, Cambridge (Massachusetts), The MIT Press, 1994; véase también Kenneth Arrow, «Methodological individualism and social knowledge», *American Economic Review: Papers and Proceedings*, 84(2), 1994, págs. 1-9 [«Individualismo metodológico y conocimiento social», *Revista Asturiana de Economía*, 22, 2001, págs. 173-186]. Me he referido en forma más sistemática a los proverbios en «Sagesse et science: le rôle des proverbes dans la connaissance de l'homme et de la société», en Jean Baechler (comp.), *L'Acteur et ses raisons: mélanges en l'honneur de Raymond Boudon*, París, Presses Universitaires de France, 2000, págs. 351-362. La idea de la «psicología de la tiranía» procede de John Roemer, «Rationalizing revolutionary ideology», *Econometrica*, 53(1), 1985, págs. 85-108. El estudio de los sujetos a quienes se prometían discos fue realizado por Jack Brehm y otros, «The attractiveness of an eliminated choice alternative», *Journal of Experimental Social Psychology*, 2, 1966, págs. 301-313. Se encontrará una introducción general a la teoría de la reactancia en Robert A. Wicklund, *Freedom and Reactance*, Nueva York, Wiley, 1974. Reid Hastie y Robyn Dawes, *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*, Thousand Oaks (California), Sage, 2001, Capítulo 3, exponen su escepticismo sobre la interacción que induce efectos de inversión. La historia de las correlaciones significativas al 5% a la que me refiero en una nota al pie figura en Robert R. McCrae y Paul T. Costa, «The paradox of parental influence», en Carlo Perris, Willem A. Arrindell y Martin Eisemann (comps.), *Parenting and Psychopathology*, Nueva York, Wiley, 1994, págs. 113-114. El ejemplo del

impacto del mal tiempo sobre los agentes de bolsa, también mencionado en una nota al pie, está tomado de Peter Kennedy, «Oh, no! I got the wrong sign! What should I do?», *Journal of Economic Education*, 36(1), 2005, págs. 77-92, que también contiene útiles comentarios sobre los costes (¡y beneficios!) de la explotación de datos en términos más generales.

Notas al pie

[23] En algunos de mis escritos anteriores he utilizado «mecanismo» para denotar lo que ahora llamo «cadena causal». En trabajos más recientes comencé a utilizar «mecanismo» en el sentido definido más adelante en este mismo capítulo. Probablemente debería haber optado por una terminología diferente, pero ya es demasiado tarde para hacerlo.

[24] Como se señala más adelante, en determinado momento la segunda explicación se propuso seriamente.

[25] Por otra parte, en el caso de algunos bienes la demanda *sube* cuando los precios aumentan. Los consumidores pueden sentirse atraídos por un bien porque es caro (el «efecto Veblen») o comprar menos pan, por ejemplo, cuando su precio cae, porque pueden permitirse reemplazarlo parcialmente por algún otro bien de calidad superior, como la carne (el «efecto Giffen»).

[26] A menudo se dice, desde luego, que la fuerza de los sentimientos altruistas hacia los otros varía inversamente con la distancia social que los separa del agente. Sin embargo, la idea de «distancia social» se asemeja más a una metáfora que a un concepto, y en todo caso «varía inversamente» es mucho menos preciso que «varía inversamente con el cuadrado de la distancia».

[27] Sin embargo, como veremos en el Capítulo 10, los proverbios no siempre son sabios.

[28] Si bien este proverbio no enuncia un par de proposiciones contrarias, podríamos imaginar enunciados proverbiales en el sentido de que «somos muy crédulos con lo que esperamos» y «somos muy incrédulos con lo que esperamos». De hecho, a menudo nos inclinamos a no creer en lo que esperamos porque hacemos una hipercorrección para evitar el pensamiento desiderativo.

[29] La idea podría definirse en términos más generales, para incluir tanto la incredulidad con respecto a lo que esperamos como la credulidad con respecto a lo que tememos. En la nota anterior mencioné un origen posible de la primera variedad. En cuanto a la segunda, tal vez el deseo de prepararse para lo peor tienda a fortalecer la creencia en que lo peor ha de ocurrir. El mero proceso de repasar un argumento mental podría grabarlo en la mente y elevar su estatus de remotamente posible a verosímil y hasta probable.

* Habitualmente, este proverbio, *«too many cooks make the soup too salty»*, se traduce como «muchas manos en un plato hacen mucho garabato» o «unos por otros y la casa sin barrer», pero en este caso decidimos traducirlo literalmente para ajustarnos al comentario que a continuación hace el autor. [T.]

[30] A la inversa, el recuerdo de una buena experiencia es una buena experiencia. Así, Tennyson: «Mejor es haber amado y perdido que jamás haber amado».

[31] A la inversa, el recuerdo de una buena experiencia devalúa el presente. Así, Donne: «Menos pesar da ser vil que haber sido justo».

[32] En ocasiones, el odio encontraba un blanco diferente. En tres aldeas del centro y el norte de Italia donde los alemanes se entregaron a salvajes represalias en 1944, algunos aldeanos todavía eran hostiles a los partisanos cincuenta años después, porque los veían como indirecta y hasta «verdaderamente» responsables de la masacre. Cuando A motiva a B a matar a C, los parientes y amigos de este último pueden encauzar su odio hacia el primero y no hacia el segundo. En los movimientos de resistencia que combatían a las tropas de ocupación alemanas, se observaban ambos mecanismos.

[33] Una característica especial de este ejemplo es que uno de los dos efectos antagónicos (la reactancia) es inducido por el otro (la barrera). En los otros ejemplos, una causa común (la opresión del tirano, digamos) provoca al mismo tiempo los dos efectos.

[34] En otra ilustración de la reactancia, los niños (y los adultos también) pueden rechazar todas las propuestas sobre cómo pasar el día, *incluso la que prefieren*, porque no quieren sentir que se les impone algo.

[35] Me tomo un poco de libertad con estos proverbios. En su significado literal, la expresión francesa *une brébis galeuse* se refiere a una oveja que padece una enfermedad provocada por un parásito arácnido.

[36] Además, Tocqueville afirmaba que la tiranía de la opinión mayoritaria ejercería una influencia moderadora. «Se ve a gobiernos que se esfuerzan en proteger las costumbres condenando a los autores de libros licenciosos. En los Estados Unidos, no se condena a nadie por esta clase de obras; pero nadie ha intentado escribirlas.»

[37] Hago alguna trampa para que el ejemplo cuadre, porque Publio Siro se refiere a préstamos y no a regalos. Aunque parece verosímil que tanto el préstamo como el obsequio de una gran suma de dinero puedan generar el resentimiento del receptor, probablemente lo hacen de diferente manera.

[38] Sólo «algo parecido», dado que el tamaño exacto que provoca el paso de la gratitud al resentimiento dependerá mucho del contexto. La misma observación es válida para los otros casos mencionados en el texto.

[39] Recuérdese, no obstante, que la tensión puede quedar irresuelta.

[40] Así, en un estudio longitudinal de la relación entre prácticas maternales y resultados infantiles, apenas 35 de 552 correlaciones fueron estadísticamente significantes «en el nivel $p < 0,05$» (lo cual quería decir que había una posibilidad en veinte de que existieran por azar), un hecho sólo evidente para quienes leyeran los apéndices del libro, que en su reedición fueron eliminados.

[41] La teoría tal vez sugiera que el mal tiempo deprime a los agentes de bolsa y los impulsa a vender. Sin embargo, los investigadores informan del resultado opuesto cuando el mal tiempo se define como una cobertura de nubes del 100%. Al cambiar esa definición a una cobertura nubosa por encima del 80%, el signo de la correlación se invierte mágicamente.

[42] Tengo dos experiencias personales. La primera vez que visité Nueva York, compré entradas para dos espectáculos de Broadway, uno de ellos armado en torno de la música de Fats Waller y el otro alrededor de la de Duke Ellington. Como quedaban muy pocas entradas, tuve que aceptar las que me ofrecieron, que, en ambos espectáculos, eran para la fila H, asiento 130. Esto era meramente misterioso, pero otra coincidencia pareció más significativa. Hay dos experiencias que sólo he vivido una vez. Una es la de ser invitado a una fiesta y luego olvidar que me han invitado. Otra es la de ser invitado a una fiesta y luego recibir una llamada del anfitrión, treinta minutos antes de la hora en que debo estar allí, para decirme que tiene que cancelarla a causa de una enfermedad. La coincidencia, que por un segundo me hizo pensar que alguien velaba por mí, es que se trataba de la misma fiesta.

[43] Hay por lo tanto tres problemas en lo concerniente al uso de la correlación como una guía para la causalidad. Primero, la correlación puede surgir puramente por azar y carecer de una interpretación causal. Segundo, la correlación tal vez tenga una interpretación causal indirecta si dos fenómenos correlacionados son efectos comunes de un «tercer factor». Tercero, la dirección de la causalidad podría ser ambigua.

[44] Uno podría rodear o al menos mitigar este problema mediante el ejercicio de la autolimitación. Si contamos con un conjunto de datos lo bastante grande, podemos en primer lugar concentrarnos en una muestra representativa e ignorar el resto. Una vez hecho todo lo posible para explicar el subconjunto de observaciones, podemos aplicar la explicación a la totalidad de los datos y ver si sigue siendo válida. De serlo, habrá menos probabilidades de que sea espuria. Otra manera de preservar la honestidad de los investigadores consistiría en que las revistas se negaran a considerar los artículos presentados para su publicación a menos que las hipótesis por verificar, así como los procedimientos para verificarlas, se entregaran al editor con una anticipación de (digamos) dos años. Ketil Bruun, investigador en el tema del alcoholismo, ha adoptado esta práctica.

Capítulo 3
Interpretación

Interpretación y explicación

En muchos escritos sobre las humanidades se ha hecho hincapié en la *interpretación* y no en la explicación. En la tradición alemana a menudo se trazó un contraste entre las «ciencias del espíritu» (*Geisteswissenschaften*) y las ciencias naturales (*Naturwissenschaften*). En las primeras, nos dicen, el procedimiento pertinente es el de la interpretación o «comprensión» (*Verstehen*). Para las segundas, el lenguaje adecuado es el de la explicación (*Erklären*). Max Weber escribió, por ejemplo, que las ciencias naturales no aspiran a «comprender» el comportamiento de las células.

Quizá nos preguntemos entonces si las ciencias sociales se apoyan en la comprensión o la explicación. Creo que la pregunta está mal planteada. A mi juicio, interpretar *es* explicar. La interpretación no es otra cosa que un caso especial del método hipotético deductivo (Capítulo 1). Los estudiosos de las humanidades no pueden, por ejemplo, utilizar la «empatía» como un atajo privilegiado hacia la interpretación del comportamiento, dado que la comprensión empática de un investigador puede diferir de la de otro. Para decidir entre interpretaciones antagónicas, tienen que confrontar estas corazonadas o hipótesis interpretativas (pues eso es lo que son) con la *experiencia*. Como he sostenido en el Capítulo 1, la experiencia incluye no sólo los hechos que tratamos de entender, sino también los *nuevos hechos* que, en otras circunstancias, no habríamos pensado en investigar.[45]

La interpretación se dirige a las acciones humanas y su producto, como las obras de arte. En el Capítulo 14 abordo el problema de la interpretación de obras literarias, y más específicamente de obras en las cuales necesitamos comprender tanto las acciones de los personajes como las elecciones del autor. Al tratar de entender otras obras literarias, así como el «arte sin palabras» de la pintura, la escultura o la música instrumental, este problema de dos niveles no se plantea. Sin embargo, también en esas formas artísticas las elecciones del artista, en principio, se prestan virtualmente al mismo análisis que propondré para las decisiones autorales. Los artistas eligen en función de algún criterio de «perfeccionamiento» que ni ellos ni nosotros somos capaces, tal vez, de formular de manera explícita, pero que se revela en la práctica cuando desechan un borrador, un bosquejo o una grabación en favor de otro. No obstante, la relación entre ese criterio de perfeccionamiento y la psicología humana es más complicada y peor entendida en las artes sin palabras que en la ficción (clásica). No intentaré ocuparme de ellas.

Racionalidad e inteligibilidad

El resto de este capítulo, por consiguiente, se dedicará a la interpretación de la *acción*. La interpretación de una acción nos exige explicarla desde el punto de vista de las creencias y deseos (motivaciones) precedentes del agente. Por otra parte, debemos explicar esos mismos estados mentales de manera de poder otorgarles un sentido, situándolos dentro de la totalidad del complejo de deseos y creencias. Un deseo o una creencia aislados que no tiene el tipo normal de solidaridad con otros estados mentales es un mero hecho en bruto que tal vez nos permita explicar el comportamiento, pero no entenderlo.

Un modo paradigmático de explicar la acción consiste en demostrar que se realizó porque era *racional* (Capítulo 11). Para hacerlo, no basta con mostrar que redundaba en buenas consecuencias para el agente: es preciso entenderla como óptima desde el punto de vista de éste. Es un hecho, por ejemplo, que si la gente atribuye un alto valor a las consecuencias futuras del comportamiento presente, es decir, tiene un bajo índice de descuento temporal (Capítulo 6), su vida mejora. También es verosímil que una educación superior configure las preferencias temporales en esa dirección. Pese a ello, estas dos premisas no equivalen a una explicación basada en la

elección racional de por qué la gente decide educarse. Para que una explicación levante vuelo, tendríamos que demostrar que la gente tiene las creencias requeridas acerca de la influencia de la educación sobre la aptitud de postergar la satisfacción, y que está subjetivamente motivada para adquirir esa aptitud.[46] El intento de explicar la elección por sus consecuencias beneficiosas es una forma de «funcionalismo de la elección racional» (una combinación de los dos enfoques contra los cuales advertí en la Introducción) que no arroja luz alguna sobre el significado del comportamiento.

Si la conducta es racional, también es ipso facto inteligible. Sin embargo, también puede serlo el comportamiento irracional. Distinguiré entre tres variedades de comportamiento inteligible pero irracional y las contrastaré con algunos casos de comportamiento ininteligible.

La primera surge cuando la maquinaria de la toma de decisiones (véase la Figura 11.1) queda *truncada* de una manera u otra. En virtud de su singular urgencia, una emoción fuerte puede impedir al agente «mirar a su alrededor» (esto es, recoger información) antes de actuar. En vez de adoptar una estrategia de espera similar a la del general romano Fabio Cunctator («el Prudente»), el agente se precipita a la acción sin tomarse el tiempo de considerar las consecuencias. Otra forma de truncamiento se da con la debilidad de la voluntad, tradicionalmente entendida como el hecho de actuar a sabiendas de que cometemos un error (Capítulo 6). La persona que ha decidido dejar de fumar pero acepta el ofrecimiento de un cigarrillo actúa por una razón, a saber, el deseo de fumar. Sin embargo, para que una acción sea racional, tiene que ser óptima a la luz de la totalidad de las razones, no de una sola de ellas. Tendré ocasión, de todos modos, de cuestionar esta concepción de la debilidad de la voluntad.

Una segunda variedad surge en la *evitación* de la maquinaria de la decisión que se produce cuando los deseos del agente influyen en la formación de creencias. El pensamiento desiderativo, por ejemplo, es irracional pero plenamente inteligible. Una forma más sutil de formación de creencias motivadas aparece cuando el agente deja de recoger información en el momento en que las pruebas recolectadas hasta entonces respaldan la creencia que él querría cierta.[47] Estas formas de formación de creencias motivadas son, a su manera, procesos de optimización: maximizan el placer que el agente extrae de sus creencias sobre el mundo, y no el placer que puede esperar de sus encuentros con éste.

Una tercera variedad es lo que podríamos llamar *cruce de cables* en la maquinaria de la decisión. Es fácil entender por qué la mente podría consagrarse a una reducción de la disonancia cognitiva (una de cuyas variedades es el pensamiento desiderativo), pero ¿por qué debe buscar también la *producción* de disonancia? La idea, mencionada en el Capítulo 2, de que creemos con facilidad lo que tememos, es un ejemplo. ¿Por qué el temor a un mal resultado habría de llevarnos a ver éste como más probable de lo que nuestras pruebas lo justifican? Si la creencia no cuenta con el respaldo de las pruebas ni de nuestros deseos, ¿por qué adoptarla? Es evidente que no hay optimización de nada. En cierto sentido, ese comportamiento es más difícil de entender que las acciones originadas en el truncamiento y la evitación, dado que *en él no hay nada* para el agente, ninguna meta parcial o de corto plazo que el comportamiento satisfaga. Pese a ello, es inteligible (tal y como yo entiendo esa idea) porque surge del sistema de creencias y deseos del agente.

Entre las acciones que eluden la interpretación se cuentan las causadas por compulsiones y obsesiones, el comportamiento fóbico, las automutilaciones, la anorexia y cosas por el estilo. Está claro que esos comportamientos tienen el efecto, que explica por qué existen, de aliviar la angustia que el agente siente si no los exhibe. No obstante, el hecho de lavarse las manos cincuenta veces al día o subir cincuenta pisos por la escalera para no usar el elevador no es lo mismo que tomar un tranquilizante. Tomar un Valium puede ser tan racional e inteligible como tomar una aspirina, pero el comportamiento compulsivo y fóbico es ininteligible porque no forma parte de un *sistema* interconectado de creencias y deseos. Otro ejemplo, debido a John Rawls: nos costaría entender el comportamiento de alguien que dedicara su tiempo a contar briznas de hierba, de no ser que esta actividad estuviera ligada a algún otro objetivo, como el de ganar una apuesta.

El pensamiento desiderativo es inteligible, como lo es el pensamiento contradesiderativo. La creencia de un individuo perturbado en que el dentista del edificio de al lado dirige rayos X contra él para destruirle la mente no lo es. En contraste, las creencias paranoicas en política son inteligibles porque tienen sus raíces en los deseos del agente. Una persona muy antisemita tiene motivos para abrigar creencias absurdas sobre la naturaleza omnipotente y maligna de los judíos (véase el Capítulo 7). No es que quiera que éstos *tengan* esas características, sino que está motivada a

creer que las tienen porque la creencia puede racionalizar su impulso de eliminarlos. Aun las creencias contradictorias pueden ser inteligibles. En diferentes oportunidades, un antisemita puede calificar a los judíos de «gusanos» *y* aseverar su omnipotencia. La misma gente que dice que «los judíos siempre tratan de entrar donde no los quieren» cree también que «los judíos son exclusivistas, siempre se juntan entre ellos». Un mismo musulmán puede afirmar que el Mossad, el servicio de inteligencia israelí, estuvo detrás de los ataques al World Trade Center el 11 de septiembre de 2001, *y* enorgullecerse por ese acontecimiento.

Entender las guerras civiles

Querría mencionar dos ejemplos extensos para dar carnadura a las ideas de las creencias y deseos inteligibles, ambos tomados de estudios de guerras civiles pasadas y presentes. Recurriré luego a los mismos estudios y algunos otros para abordar la cuestión hermenéutica fundamental de cómo podemos atribuir o establecer motivaciones y creencias.

Consideremos en primer lugar la creencia en la predestinación, que fue uno de los principales problemas que dividieron a calvinistas y católicos en las guerras de religión. En su origen se encontraba la intensa angustia religiosa experimentada por muchos creyentes en épocas previas a la Reforma, debido a la incertidumbre en torno de la salvación propia. ¿Cómo podía estar uno seguro –o podía estarlo alguna vez– de haber hecho lo suficiente para alcanzarla? En una mirada retrospectiva a sus primeros años, Calvino escribió en 1539 que aun cuando había cumplido las exigencias de la Iglesia de confesar sus pecados y borrarlos de la memoria de Dios por medio de las buenas obras y la penitencia, «estaba muy lejos de la certidumbre y la tranquilidad de conciencia. Puesto que cada vez que me sumergía en mí mismo o elevaba mi corazón hacia Ti, me asaltaba tan extremo horror que ni las purgaciones ni las exculpaciones podían curarme».

Calvino se liberó de la angustia cuando abandonó la concepción de Dios como inmanente en el mundo, una presencia opresiva y amenazadora, y lo concibió como absolutamente trascendente. Esta idea tenía un vínculo crucial con la doctrina de la doble predestinación: como Dios había decidido desde la eternidad quiénes se salvarían y quiénes se condenarían, *no había nada que uno pudiera hacer* por su propia salvación y, por tanto,

tampoco motivos para preocuparse por no haber hecho lo suficiente. El problema interpretativo clave concierne al vínculo entre esta creencia en la predestinación y la liberación de la angustia. A priori, este efecto de la doctrina parece ininteligible. Calvino enseñaba que los elegidos eran una pequeña minoría, que oscilaba (en diferentes exposiciones) de uno en cien a uno en cinco. ¿Qué podía suscitar más angustia que la creencia en que uno tenía muchas probabilidades de contarse entre los condenados, sin que pudiera hacer nada para evitar arder por toda la eternidad en el infierno? ¿La conversión del catolicismo al calvinismo no significaba, literalmente, pasar de la sartén al fuego?

Es probable que la respuesta deba encontrarse en los conceptos esbozados por primera vez por Max Weber. Dada su creencia en la predestinación, los calvinistas no podían sostener que un esfuerzo racional y sistemático les traería la salvación, pero sí podían alegar, y alegaron, que les daría la *certeza* subjetiva de la salvación. El propio Calvino escribió que «la vocación de los elegidos es como una demostración y un testimonio de su elección». Y parece, de hecho, que la conversión al calvinismo eliminó efectivamente la incertidumbre con respecto a la salvación. Vuelvo a esta forma de «pensamiento mágico» en el Capítulo 7. Aquí quiero limitarme a destacar que los mecanismos gemelos del pensamiento desiderativo y el pensamiento mágico prestan inteligibilidad a la creencia en la predestinación.

Tomemos en consideración ahora la inteligibilidad de las motivaciones. ¿Por qué los jóvenes palestinos están dispuestos a dar la vida en misiones suicidas? Su principal motivación (procurarse o defender un hogar nacional) no es difícil de entender.[48] Esta causa puede ser tan apremiante como lo fue la defensa de la democracia en la lucha contra Hitler. Lo que acaso parezca enigmático es la *fuerza* de la motivación. Para hacerla inteligible, hacen falta algunos factores causales adicionales. Examinaré media docena y me inclinaré por uno de ellos.

Con anterioridad al 11 de septiembre, existía la difundida creencia de que el atacante suicida típico en Medio Oriente era un hombre joven soltero y desempleado, tal vez sexualmente famélico, para quien un movimiento religioso podía llenar un vacío que en otras circunstancias habrían ocupado la familia y el trabajo. Luego, de la noche a la mañana, tras el ataque al World Trade Center, los expertos en terrorismo decidieron que tenían que «reescribir los libros». Sin embargo, aun antes de ese acontecimiento, la

frecuente aunque fluctuante intervención de mujeres atacantes suicidas debería haber llevado a los estudiosos a cuestionar ese estereotipo. En la segunda *intifada*, el uso de mujeres suicidas, algunas de ellas madres o muy instruidas, ha sido aún más sorprendente.

Los factores de la pobreza y el analfabetismo, muchas veces mencionados, también parecen tener limitada eficacia causal, al menos como rasgos individuales de los atacantes suicidas. Entre los suicidas palestinos, el ingreso y la educación suelen ser, de hecho, más elevados que en la población general. Las explicaciones planteadas desde el punto de vista de la pobreza también son insatisfactorias, porque no resulta claro cómo habría generado ésta la motivación requerida. Según una concepción corriente, las ganancias de volarse deben contrapesarse con el coste de hacerlo: la propia vida. Si no se atribuye a ésta mucho valor, el coste es menor. De acuerdo con este enfoque, una vida en la desventura y la pobreza vale tan poco para el individuo, que los costes del suicidio resultan desdeñables. Soy escéptico con respecto a este argumento, pues me parece que los pobres consideran su vida tan digna de vivirse como cualquier otra persona. Un dato psicológico bastante bien establecido es que la gente ajusta sus aspiraciones a sus circunstancias, con el fin de poder mantener un nivel más o menos constante de satisfacción («la noria hedonista»).

Un factor más verosímil que la privación absoluta es la *privación relativa*, es decir la brecha entre las expectativas y la realidad experimentada por los muchos palestinos educados que hoy carecen de toda perspectiva de conseguir un empleo decente. La movilidad social descendente podría tener el mismo efecto. Sin embargo, las características más relevantes parecen ser las sensaciones permanentes de *inferioridad* y *resentimiento*. La primera de estas emociones se basa en la *comparación* entre uno mismo y otros, y la segunda, en la *interacción* entre uno mismo y los demás. Hablando en general, las emociones basadas en la interacción son más poderosas que las basadas en la comparación. Muchos autores que se refieren a los atacantes suicidas palestinos hacen hincapié en el intenso resentimiento causado por las humillaciones cotidianas que deben vivir en su interacción con las fuerzas israelíes. Más allá de las degradantes inspecciones y controles a los que son sometidos los palestinos, está también su conciencia de que muchos israelíes creen que todos los árabes son «perezosos, cobardes y crueles», como me dijo un taxista de Jerusalén hace unos veinte años.

Si esta descripción es correcta, el fuerte resentimiento contra quienes ocupan actualmente la deseada patria nos permite *entender* la disposición a morir de los atacantes suicidas palestinos. El deseo de combatir a los israelíes extrae su fuerza del hecho de estar inmerso en un complejo motivacional más grande. Hay, con todo, un punto de vista alternativo. Los atacantes suicidas son manejados con rienda corta por quienes los manipulan, que se apresuran a ejercer presiones adicionales en caso de que la motivación primaria falle al acercarse el momento de la acción. En Irak, un aspirante a atacante suicida, que fue capturado y desarmado porque estaba notoriamente nervioso, dijo que antes de su misión había estado encerrado tres días en una habitación con un mulá que le hablaba del Paraíso y lo alimentaba con «una sopa especial que lo fortalecía». El estado mental que desencadena efectivamente el acto de detonar la bomba puede ser, por lo tanto, efímero y semejante a un artefacto, y no un rasgo estable de la persona. Si bien expresiones como «lavado de cerebro» e «hipnosis» tal vez sean demasiado fuertes, hay pruebas de que algunos de los atacantes se encontraban en un estado similar al trance en los minutos previos a su muerte. Cuando, como en tales casos, una intención queda aislada del sistema global de deseos y creencias de la persona, ninguna interpretación es posible. En cambio, el comportamiento de los manipuladores y, más en general, de los organizadores de la misión, puede ser desde luego objeto de una interpretación.

Un dilema hermenéutico

Está muy bien afirmar que el comportamiento debe explicarse en función de los estados mentales antecedentes, deseos y creencias, que lo causan, pero ¿cómo establecemos esas causas previas? So pena de caer en la circularidad, no podemos utilizar como prueba el comportamiento mismo. Debemos buscar otras pruebas, como declaraciones del agente acerca de su motivación, la coherencia de su comportamiento no verbal con esas declaraciones, los motivos atribuidos a él por otros y la coherencia del comportamiento no verbal *de éstos* con esas atribuciones. ¿Cómo podemos, empero, excluir la posibilidad de que esas formas verbales y no verbales de comportamiento hayan sido elegidas de manera deliberada para llevar a una audiencia a creer, falsamente, en la existencia de una motivación en particular? Las profesiones y alegaciones mismas de motivaciones pueden

estar motivadas. La cuestión es central en la toma colectiva de decisiones. Como sostengo en el Capítulo 25, todos los métodos para reunir las preferencias individuales en una decisión social generan incentivos para que los participantes en el proceso distorsionen sus preferencias.

Examinemos, como ejemplo, los motivos de líderes y seguidores en las guerras civiles. Las partes han profesado, o sus adversarios les han atribuido, uno de tres motivos: *religión*, *poder* y *dinero*. Quienes profesan motivos religiosos son acusados a menudo de utilizarlos como un disfraz de sus verdaderos móviles, sean éstos políticos o pecuniarios. Durante las guerras francesas de religión (1562-1598), los bandos en pugna se acusaban constantemente uno a otro de usar la religión como pretexto para sus objetivos políticos e incluso crematísticos. Las acusaciones tenían cierto fundamento. Enrique de Navarra (luego Enrique IV) se convirtió seis veces en su vida, y la última conversión, en 1593, fue objeto de amplias sospechas de oportunismo. Su padre, Antonio de Borbón, ya había señalado a las claras que ofrecía su fe en venta al mejor postor. Y solía acompañar a la reina regente a misa y a su esposa protestante en la comunión. En su lecho de muerte, buscó el consuelo de ambas religiones. Un importante reformador, el cardenal de Châtillon, se casó luego de su conversión, pero conservó tanto su título de cardenal como los ingresos procedentes de su diócesis. Otro prelado, Antoine Caracciolo, obispo de Troyes, también quiso combinar un ministerio protestante con las rentas de su obispado. Un católico de nota como Enrique, duque de Guisa, estaba perfectamente dispuesto a concertar una alianza con los calvinistas contra el rey Enrique III.

También en el mundo contemporáneo la religión se utiliza a veces como pretexto para la política, y la política como pretexto para el dinero. Los objetivos de los insurgentes chechenos y de algunas organizaciones palestinas, sobre todo Al Fatah, fueron en su origen exclusivamente políticos. Cuando adoptaron un manto religioso, lo hicieron en gran medida para conquistar un mayor número de partidarios. En Palestina, la rivalidad con Hamas, un grupo incuestionablemente religioso, hizo de esa actitud una necesidad para la supervivencia de Al Fatah como organización. En las Filipinas, el grupo terrorista Abu Sayyaf ha usado la demanda de un Estado islámico independiente como un pretexto para secuestrar y obtener enormes rescates. En Colombia, no existe la certeza de que las Fuerzas Armadas Revolucionarias de Colombia (FARC) mantengan su motivación original de

luchar contra la injusticia social y no hayan degenerado a estas alturas en un grupo mafioso. En todos estos casos, como en las guerras francesas de religión, la atribución de motivos suele estar colmada de incertidumbre. Tal vez sea difícil saber, en especial, si hay completa coincidencia entre los motivos de los líderes y los de sus seguidores.

Hay muchas razones por las cuales la gente puede llegar a querer distorsionar sus motivaciones y las de sus adversarios. Por un lado, cada sociedad tiene una jerarquía normativa de motivaciones (Capítulo 4) que induce en sus miembros el deseo de mostrarse animados por la nobleza y no por la mezquindad de sus motivos, y achacar un móvil de baja estofa al oponente. En las guerras francesas de religión, al igual que en la guerra civil inglesa, cada bando se presentaba como si estuviera imbuido de motivos religiosos, y sostenía que la mera sed de poder movía al otro. Por otro lado, si uno puede hacer que los demás den crédito a su profesión de una motivación específica, quizá le resulte más fácil alcanzar sus metas. Como la imagen de un terrorista puede ser más intimidante que la de un delincuente común, es probable que los secuestradores mercenarios incrementen sus posibilidades de lograr concesiones si esgrimen las banderas de una causa. En Colombia, muchos secuestros son cometidos por delincuentes comunes que se declaran pertenecientes a un grupo guerrillero para tratar de provocar miedo en la familia de sus víctimas. Los secuestros son más alarmantes si se cree que los terroristas están dispuestos a tomar medidas drásticas en caso de que algo resulte mal, y son menos propensos a negociar plazos o regatear por el dinero pedido como rescate. Si no pueden obtener lo que exigen, al menos pueden «fijar posición» matando a sus víctimas.

El problema del sesgo autorreferencial en las declaraciones sobre las intenciones de los agentes sociales es grave, pero no insuperable. Una manera sencilla de eludirlo podría consistir en considerar los *intereses objetivos* del agente y suponer que, en ausencia de pruebas sólidas en contrario, coinciden con su motivación subjetiva, con prescindencia de lo que el agente diga de esta última. Otra alternativa sería identificar las *consecuencias reales* de su acción y suponer que, en ausencia de pruebas sólidas en contrario, son las que el agente quería provocar. (Una y otra idea podrían aplicarse a la elección de una educación superior antes examinada.) El hecho de que existan estos *dos* procedimientos para trasladar el peso de la prueba sugiere, sin embargo, que ninguno de ellos es aceptable. Tanto los

intereses objetivos como las consecuencias reales pueden señalar hipótesis útiles sobre los motivos subjetivos, pero ni unos ni otras tienen una presunción en su favor.

Los historiadores y los especialistas en ciencias sociales han elaborado otros métodos de manejar el problema que, sobre todo cuando se combinan, pueden producir conclusiones razonablemente certeras. Una técnica consiste en ir más allá de las declaraciones hechas ante una audiencia y buscar las menos probablemente motivadas por un deseo de distorsión. Las cartas, los diarios íntimos, las conversaciones divulgadas, los borradores y otras cosas por el estilo pueden ser fuentes invalorables. Sabemos, por las cartas que escribieron a sus esposas, que algunos delegados a la Asamblea Constituyente francesa de 1789 habían votado contra el bicameralismo y el veto real porque creían que su vida podía correr peligro si votaban de otra manera. En la asamblea, justificaron su decisión sobre la base del interés público. Al tratar de indagar en las motivaciones que impulsaron la masacre de la noche de San Bartolomé en 1572, los historiadores han comprobado la utilidad de ir más allá de los relatos poco imparciales de los participantes y apoyarse en los informes de diplomáticos extranjeros que estaban interesados en comprender con claridad la situación. En la Inglaterra decimonónica, las declaraciones hechas en el lecho de muerte quedaban eximidas de las reglas habituales sobre las pruebas de oídas. El primer borrador de un documento puede decir más acerca de las creencias y motivos del autor que una ulterior versión publicada. Es instructivo, por ejemplo, comparar los borradores de *La guerra civil en Francia*, de Marx, o de su carta a Vera Zasulich, con las versiones oficiales.

También puede haber un pronunciado contraste entre lo que los actores dicen en público y lo que dicen a puertas cerradas. Aunque la publicación de los debates de la Asamblea Constituyente francesa de 1789-1791 es fascinante de principio a fin, dos factores conspiran para hacerlos poco confiables como pruebas de estados mentales. Por un lado, la publicidad del marco forzaba a los delegados a valerse únicamente de argumentos fundados en el interés público; la abierta manifestación de intereses de grupo era inadmisible. Por otro, la presencia de un millar de delegados e igual cantidad de público en las galerías alentaba la vanidad de los oradores. En ambos aspectos, la Convención Federal norteamericana fue un ámbito más favorable para la sinceridad. Como el número de delegados era pequeño (cincuenta y cinco, contra mil doscientos en París) y las sesiones

se realizaban bajo el velo del secreto, podía haber y había negociaciones basadas en distintos intereses. Al mismo tiempo, y como Madison escribió muchos años después:

> Si en un principio los representantes se hubieran comprometido públicamente, habrían supuesto luego que la coherencia les exigía mantenerse en sus trece, en tanto que la discusión secreta hacía que sólo se sintieran obligados a insistir en sus opiniones mientras estuvieran convencidos de su propiedad y su verdad, por lo cual eran receptivos a la solidez de la argumentación.

El temor a las revelaciones futuras tampoco desalentaba los debates, dado que el secreto debía extenderse de manera indefinida y, de hecho, sólo fue roto con la publicación de las notas de Madison muchas décadas después. Cuando la sinceridad no tiene costes, las razones estratégicas de la distorsión pierden fuerza.

Los especialistas en ciencias sociales también pueden eliminar el coste de la sinceridad creando un velo artificial de ignorancia. Supongamos que un investigador quiere estudiar la relación entre la orientación sexual y algunas otras variables de interés. Quizá sea difícil inducir a los sujetos a responder en forma veraz a la pregunta de si han tenido alguna vez experiencias sexuales con personas del mismo sexo, aun cuando se les asegure el anonimato de sus respuestas. Para eludir este problema, el investigador puede indicarles que respondan honestamente en caso de que hayan tenido alguna experiencia de ese tipo y, si nunca la han tenido, que tiren una moneda para decidir si la respuesta es sí o no. Si cumplen esa indicación, y no tienen razones para no hacerlo, y la muestra es lo bastante amplia, los datos serán tan válidos como si todos hubieran contestado sinceramente.

Otra técnica para verificar si el comportamiento no verbal de los agentes es coherente con su motivación manifiesta consiste en preguntarse si predican con el ejemplo. Cuando en 2003 el gobierno de Bush mencionó su certeza de que Saddam Hussein tenía armas de destrucción masiva como la principal razón para invadir Irak, ¿también implementó las medidas imprescindibles para proteger a los soldados estadounidenses de esa amenaza? Algunos patrones de conducta pueden revelar la verdadera motivación de los secuestradores. En 1996, en Costa Rica, unos secuestradores (principalmente ex contras de Nicaragua) exigieron, para liberar a sus víctimas, un millón de dólares de rescate, además de garantías laborales para los trabajadores, una reducción del precio de los alimentos, un aumento del salario mínimo costarricense y la liberación de varios de sus

camaradas encarcelados. Cuando se les ofrecieron doscientos mil dólares, quedaron satisfechos y no insistieron en las demandas políticas, un hecho que convenció a las autoridades de que su postura rebelde y a lo Robin Hood era un ardid y de que su objetivo siempre había sido el dinero. Otro ejemplo es el comportamiento de los aristócratas emigrados a Londres durante la Revolución francesa. En ese semillero de rumores sobre la inminente restauración de la monarquía y la competencia para ver quién era más realista, era vital para cada uno transmitir su disposición a ponerse al servicio de la contrarrevolución. Las seguridades verbales no bastaban. Cualquier persona que alquilara un apartamento por más de un mes era mal vista; era mejor alquilarlos por semana, para no dejar dudas de que uno estaba listo para regresar a Francia ni bien la contrarrevolución lo convocara.

No sólo los contemporáneos, sino también los historiadores utilizan de manera habitual esos indicadores de la conducta para juzgar la sinceridad de profesiones públicas de lealtad. Hacia el final de la Segunda Guerra Mundial, por ejemplo, en la Francia ocupada había un marcado grado de escepticismo acerca de las perspectivas de una victoria alemana. Tal vez no fuera seguro expresar la actitud, pero ésta se reflejaba en el comportamiento. La proporción de estudiantes secundarios que escogían el alemán como lengua extranjera (o cuyos padres lo escogían para ellos) se duplicó entre 1939 y 1942 y de allí en más cayó rápidamente. Muchas editoriales que se habían precipitado a firmar contratos para la traducción de libros alemanes decidieron no hacer uso de la opción.

Jueces y jurados proceden con frecuencia de la misma manera. A veces preguntan: «¿El acusado tenía un motivo para hacer X?», con la esperanza de que una respuesta les ayude a decidir si efectivamente aquél ha hecho X. En este caso, «tener un motivo» es una idea objetiva, a saber, si el hecho de hacer X aportaba algún beneficio al acusado. En otros casos, más pertinentes aquí, se establece que el acusado hizo X y la pregunta es: «¿Cuál fue su motivo para hacerlo?» Para resolver si un asesinato fue un crimen pasional o un acto cometido a sangre fría, los jueces y los jurados no buscan principalmente los beneficios objetivos, sino que tratan de establecer el estado de ánimo subjetivo del acusado. Si éste afirma haber actuado en un arranque de ira o celos y luego se demuestra que había comprado el arma asesina con antelación o que se tomó su tiempo para matar,[49] su credibilidad se ve debilitada.

Individualmente consideradas, cada una de estas técnicas puede fracasar. Un diputado podría no estar dispuesto a admitirle a su mujer que temía por su vida, o podría afirmar que tenía miedo con el fin de ocultar un motivo menos respetable (por ejemplo, aceptar un soborno). En la India decimonónica, las declaraciones en el lecho de muerte eran vistas como poco confiables, dado que la gente a veces utilizaba sus momentos de agonía para perjudicar a sus enemigos. En el ejemplo de los emigrados, *tanto* los verdaderos creyentes *como* los incrédulos tendrían motivos para tomar alquileres semanales, los primeros para facilitar su regreso a Francia cuando llegara la hora y los segundos para eludir la acusación de derrotistas. Sin embargo, la capacidad de la gente de tejer la enmarañada red de engaños sin revelar sus verdaderos motivos tiene límites. La hipocresía, dijo Somerset Maugham, es una profesión de tiempo completo. Aun Tartufo resbala al final. Para defender la sinceridad y las creencias religiosas de Enrique IV, su biógrafo no sólo menciona las pruebas positivas de «numerosos episodios en que su espíritu religioso se manifestó sin ninguna intención propagandística», sino que también argumenta: «Si hubiese habido alguna hipocresía, habría mostrado sus cuernos en tal o cual ocasión placentera». En ese mismo sentido, podemos citar a Montaigne:

> Aquellos que alegan generalmente contra lo que yo profeso, diciendo que lo que llamo franqueza, sencillez y naturalidad en mis costumbres, es habilidad y sutileza y más bien prudencia que bondad, esfuerzo que espontaneidad, buen juicio que ventura, concédenme más honor del que me quitan. Mas es verdad que creen mi sutileza demasiado sutil, y a quien me siga y me espíe de cerca concederé que ha ganado si no confiesa que no hay regla alguna en su escuela que pueda imitar ese movimiento natural y mantener una apariencia de libertad y de licencia tan igual e inflexible por esos caminos tan tortuosos y diversos, y que toda su atención e inteligencia no sabrían guiarle.

Si bien es posible que los beneficios de la distorsión sean considerables, los costes pueden ser prohibitivos. En alguna medida, la profesión instrumental de motivos es autolimitante. Dado que cualquier motivo está inmerso en una vasta red de otros motivos y creencias, la cantidad de ajustes que deben hacerse para sostener la hipocresía puede ser agobiante. Una sola nota falsa tal vez baste para echar abajo toda la construcción. Muchos proverbios dan testimonio de la irreversibilidad del derrumbe de la verdad. Si bien hay que restringir seriamente la creencia popular «mentiroso una vez, mentiroso siempre» (Capítulo 10), la creencia *irrestricta* tiene de hecho amplia circulación y sirve hasta cierto punto como disuasivo de la

mentira. Por esta razón, entre otras, Descartes quizás haya acertado al decir que «la mayor sutileza de todas es no hacer nunca uso de la sutileza».

* * *

NOTA BIBLIOGRÁFICA

El debate sobre *Erklären* contra *Verstehen* se aborda en la tercera parte de Michael Martin y Lee C. McIntyre (comps.), *Readings in the Philosophy of Social Science*, Cambridge (Massachusetts), The MIT Press, 1994. Dagfinn Føllesdal sostiene una posición cercana a la mía en el capítulo «Hermeneutics and the hypothetico-deductive method», págs. 233-245. La cita de Max Weber pertenece a su artículo «The interpretive understanding of social action», incluido en May Brodbeck (comp.), *Readings in the Philosophy of the Social Sciences*, Londres, Macmillan, 1969, pág. 33. La incoherencia de las actitudes antisemitas es brevemente abordada en Joseph Telushkin, *Jewish Humor: What the Best Jewish Jokes Say about the Jews*, Nueva York, V. Morrow, 1992, que también es mi fuente para otras observaciones sobre los judíos y de ellos mismos acerca de sus presuntas características. Los comentarios sobre la elección educacional son una polémica implícita contra Gary Becker y Casey Mulligan, «The endogenous determinations of time preferences», *Quarterly Journal of Economics*, 112, 1997, págs. 729-758. El valor probatorio de las confesiones en el lecho de muerte se examina en James Fitzjames Stephen, *A History of the Criminal Law of England*, Londres, Macmillan, 1883 y Búfalo (Nueva York), Hein, 1964, vol. 1, págs. 447-449. Henning Sass, «Affektdelikte», *Nervenarzt*, 54, 1983, págs. 557-572, enumera trece razones por las que la afirmación de que un crimen se ha cometido por motivos pasionales podría carecer de credibilidad. Un sobresaliente examen interpretativo de las motivaciones en las guerras francesas de religión es Denis Crouzet, *Les Guerriers de Dieu: la violence au temps des troubles de religion, vers 1525-vers 1610*, Seyssel, Champ Vallon, 1990. Se encontrarán análisis interpretativos de las motivaciones y creencias de los atacantes suicidas en Luca Ricolfi, «Palestinians, 1981-2003», Stephen Holmes, «Al-Qaeda, September 11, 2001», y Jon Elster, «Motivations and beliefs in suicide missions», artículos incluidos en Diego Gambetta (comp.), *Making Sense of Suicide Missions*, Oxford y Nueva York, Oxford University Press, 2005, págs. 77-129, 131-172 y 233-258, respectivamente. El comentario sobre la creencia religiosa

de Enrique IV pertenece a Jean-Pierre Babelon, *Henri IV*, París, Fayard, 1982, pág. 554. El escepticismo excesivo con respecto a los motivos se discute en Gerrie Mackie, «All men are liars: is democracy meaningless?», en Jon Elster (comp.), *Deliberative Democracy*, Cambridge y Nueva York, Cambridge University Press, 1998, págs. 69-96 [«Todos los hombres son mentirosos: ¿carece de sentido la democracia?», en *La democracia deliberativa*, Barcelona, Gedisa, 2000].

Notas al pie

[45] En las ciencias experimentales, «nuevos hechos» puede significar hechos que son literalmente novedosos, como cuando se expone a ratas o seres humanos a condiciones que no se dan en la naturaleza. En las humanidades y las ciencias no experimentales, «nuevo» debe tomarse en el sentido epistemológico de «previamente no sospechado» y no en el sentido ontológico de «previamente inexistente».

[46] En el Capítulo 11 sostengo que la idea de estar motivado para estar motivado por las consecuencias de largo plazo es conceptualmente incoherente, pero ésta es una cuestión independiente de la que planteo aquí.

[47] En una época de mayor inocencia estadística, Gregor Mendel, el descubridor de las leyes de la herencia genética, aparentemente se valió en sus experimentos de este método de «abandonar cuando uno está a la cabeza».

[48] Al afirmar que ésa es su principal motivación, no niego que pueda haber otras, como el deseo de gloria o fama póstumas, los beneficios materiales que tocarán a la familia del atacante suicida, la venganza por el asesinato israelí de un amigo o pariente o la presión social para ofrecerse como voluntarios en una misión. Según señalo en la Introducción a la segunda parte, soy escéptico en cuanto al poder motivacional de los beneficios religiosos que asumen la forma de un acceso privilegiado al Paraíso.

[49] En un fallo británico (*R. v. McPherson*) de 1957, lord Goddard se preguntó en forma retórica: «¿Cómo puede decirse que el apelante actuaba en un arrebato pasional cuando disparó no un tiro sino cuatro, y cada tiro implicaba abrir el arma para recargarla y sacar los cartuchos en cuatro ocasiones distintas?»

SEGUNDA PARTE

La mente

Este libro se organiza en torno del modelo de la acción basado en las creencias y deseos. Para entender de qué manera actúa e interactúa la gente, antes es preciso comprender cómo funciona su mente. En buena parte, ésta es una cuestión de introspección y psicología popular, refinada y corregida por los estudios más sistemáticos llevados a cabo por los psicólogos y, en medida creciente, por los economistas conductistas. El modelo es vital no sólo para explicar el comportamiento, sino también para asignar elogios, censuras o castigos. Por lo común, la culpa presupone dolo, intenciones y creencias. La responsabilidad estricta (la culpa exclusivamente atribuida sobre la base de las consecuencias reales de la acción) es poco habitual. De hecho, a veces consideramos culpable a alguien sobre la mera base de las intenciones, aun cuando éstas no hayan tenido consecuencias. El intento de homicidio es un crimen. «Las brujas», dijo John Donne, «creen en ocasiones que matan cuando en realidad no lo hacen y, por lo tanto, son tan culpables como si lo hicieran.» «En cuanto a las brujas», escribió Hobbes, «no creo que su brujería sea un poder real; pero sí, no obstante, que se las castigue con justicia, por su falsa creencia de que son capaces de cometer ese daño, unida a su propósito de hacerlo si pueden.»

Aunque indispensable, el modelo de las creencias y los deseos es frágil. Los métodos que utilizamos para atribuir estados mentales a otras personas no siempre producen resultados estables. Si queremos medir la altura de un edificio, no importa que lo hagamos desde el techo hacia abajo o desde el suelo hacia arriba. En la determinación de las creencias y los deseos, el resultado puede depender de esos factores irrelevantes. Considérese, por ejemplo, la idea de que la gente «maximiza la utilidad esperada» (Capítulo 11). Para darle precisión, debemos suponer que las personas tienen una idea clara y estable del *valor* atribuido a cada resultado posible de una acción, y de la *probabilidad* que asignan a la aparición de dicho resultado. Con frecuencia, ese supuesto se justifica, pero a veces no.

Tomemos en cuenta en primer lugar las creencias del agente. Cuando se trata de deducir las probabilidades subjetivas que un individuo atribuye a un suceso, se puede utilizar el siguiente procedimiento convencional. A partir de un número *p*, preguntamos a la persona si preferiría una lotería en la que ganara cierta suma de dinero con una probabilidad *p* u otra en la que ganara la misma cantidad si el suceso en cuestión ocurre.[1] Si prefiere la primera, le ofreceremos una nueva alternativa con la probabilidad ajustada hacia abajo; si prefiere la segunda, ajustamos la probabilidad hacia arriba. De seguir

procediendo de este modo, llegaremos a la larga a una probabilidad *p** tal que al individuo le sea indiferente optar por una lotería en la que gana el dinero con esa probabilidad *p** y otra en la que gana si el suceso se produce. Podemos afirmar, entonces, que la probabilidad revelada o deducida que el individuo atribuye al suceso es *p**. En principio, *p** debería ser independiente de la *p* inicial: esto es, la probabilidad deducida debería ser independiente del procedimiento de deducción. En la práctica no sucede así: una *p* más elevada induce una *p** más elevada. Este descubrimiento sugiere que, al menos hasta cierto punto, *no hay hecho concreto*, ningún estado mental estable, que sea captado por el procedimiento.[2]

Otros procedimientos son aún más frágiles. A menudo, los investigadores atribuyen probabilidades subjetivas a los agentes, con el supuesto de que, cuando conozcan poco de la situación, asignarán igual probabilidad a cada uno de los estados posibles del mundo. Se presume que la justificación de este procedimiento es el «principio de razón insuficiente»: si no tenemos fundamentos positivos para creer que un estado del mundo es más probable que otro, la lógica nos obliga a atribuir iguales probabilidades a ellos. Pero los estados del mundo pueden conceptualizarse y enumerarse de muchas maneras. Supongamos que perseguimos a un ladrón y llegamos a una encrucijada de la que parten tres caminos, dos que van cuesta arriba y uno que va cuesta abajo. Como no tenemos motivos para creer más probable que el ladrón haya seguido un camino y no otro, la probabilidad de que haya tomado el que va cuesta abajo debería ser, de acuerdo con el principio, de un tercio. Pero como tampoco tenemos razones para creer que haya ido cuesta arriba y no cuesta abajo, la misma probabilidad debería ser de un medio. En este caso, al menos, el principio de razón insuficiente es demasiado indeterminado para servir de algo en la construcción o atribución de probabilidades.

Examinemos a continuación la deducción de preferencias. En algunos experimentos se preguntó a los sujetos si comprarían diversos artículos (accesorios de computación, botellas de vino y cosas por el estilo) por una cifra en dólares igual a los dos últimos dígitos de su número de la seguridad social. Subsecuentemente, se les pidió que mencionaran el precio máximo que estaban dispuestos a pagar por el producto. Se comprobó entonces que su número de la seguridad social tenía una influencia significativa sobre la suma que aceptaban pagar por propia voluntad. Por ejemplo, los sujetos cuyos números de seguridad social se situaban en el quintil más alto estaban

dispuestos a pagar, en promedio, cincuenta y seis dólares por un teclado inalámbrico de ordenador, mientras que los situados en el quintil más bajo sólo aceptaban pagar dieciséis dólares. Aunque los procedimientos estaban diseñados para utilizar o deducir preferencias preexistentes, los resultados muestran que no había nada que deducir: ningún hecho concreto. Las cifras debían más al anclaje proporcionado por los números de la seguridad social que a ninguna preferencia «real».

También hay pruebas de que las *compensaciones* entre valores son sumamente inestables y pueden deber tanto a los artefactos procedimentales como a una realidad mental subyacente. En los experimentos, dichas compensaciones pueden ponerse de relieve sea por *elección* o por *comparación*. Puede ofrecerse a los sujetos la alternativa entre salvar muchas vidas a un alto coste por vida salvada (A) y salvar menos vidas a un coste más bajo (B). Otra opción consiste en pedirles que indiquen un coste por vida salvada tal que les resulte indiferente salvar la mayor cantidad de vidas a ese precio (alternativa C) o elegir la alternativa B. Supongamos que un sujeto determinado menciona un coste menor que el de A. Como a esta persona le da igual elegir C o B y es posible suponer que prefiere C a A (porque C salva más vidas a menor coste), debería preferir y por lo tanto elegir B en desmedro de A. De hecho, una mayoría abrumadora de los sujetos eligió para C un coste por debajo del coste de A, y pese a ello dos tercios declararon que preferirían A a B. El valor más importante, salvar vidas, tiene mayor preponderancia en la elección que en la comparación, aunque desde un punto de vista lógico ambos procedimientos deberían ser equivalentes.

Hay otras razones por las cuales no siempre debemos tomar al pie de la letra los enunciados sobre creencias y otros estados mentales. En este aspecto, las creencias religiosas son especialmente problemáticas. En la Inglaterra de comienzos del siglo XVII, un prelado como el obispo Andrewes podía al mismo tiempo afirmar que la peste era un castigo que Dios imponía a los pecadores *y* huir de Londres para refugiarse en el campo. La creencia de que, en virtud de su origen divino, los reyes franceses podían curar las escrófulas tocando a la persona enferma, estaba en notorio retroceso hacia fines del siglo XVIII, cuando la fórmula tradicional («el rey te toca; Dios te cura») fue reemplazada por un subjuntivo («el rey te toca; que Dios te cure»). La avidez con que la corte procuraba obtener pruebas documentadas

de curaciones exitosas también sugiere una creencia poco segura de sí misma.

Para referirnos a un ejemplo contemporáneo, tomemos en consideración la idea de que el comportamiento de los atacantes suicidas islámicos puede explicarse, al menos en parte, por su creencia en que hay una vida después de la muerte a la que, gracias al martirio, tendrán un acceso privilegiado. Cabe preguntarse si esa «creencia» es de la misma naturaleza que nuestra creencia en que mañana saldrá el Sol, esto es, si se la utiliza con igual confianza como premisa para la acción. No se trata de una cuestión de certeza contra probabilidad, sino de confianza contra falta de confianza. Tal vez yo tenga gran confianza en una creencia probabilista fundada en muchas situaciones pasadas, y esté dispuesto a apostar sobre su base. Es probable que la creencia en la vida después de la muerte, abrigada por la mayoría de la gente, no sea así. Antes bien, puede ser una «cuasicreencia» un tanto nebulosa, mantenida por su valor de consumo y no como una premisa para la acción. Si todos los que afirman creer en la vida después de la muerte creyeran en ella con plena certeza, o con una «probabilidad confiable», veríamos muchos martirios más de los que de hecho observamos. Aunque algunos creyentes sean de ese tipo, y los atacantes suicidas puedan reclutarse de manera desproporcionada en ese subconjunto, sospecho que, para muchos, la religión sirve como consuelo una vez que se ha tomado la decisión, y no como premisa para ésta.[3]

De manera similar, la gente puede experimentar o decir experimentar «cuasiemociones» que difieren de las emociones auténticas por carecer de implicaciones para la acción. Algunas personas que sostienen estar indignadas por la pobreza del Tercer Mundo, y a pesar de eso nunca abren su billetera, pueden disfrutar de su indignación como un bien de consumo, porque las hace pensar bien de sí mismas. Del mismo modo, el visible gozo de muchos que pretendían sentir pena (o «cuasipena») tras la muerte de la princesa Diana era incompatible con el horrible sentimiento del auténtico pesar. El término apropiado para referirse a sus sentimientos es, creo, «sentimentalismo». (El alemán *Schwärmerei* es aún más adecuado.) Oscar Wilde definió al sentimental como la persona «que desea darse el lujo de una emoción sin pagar por ella». Ya el pago asuma la forma de donaciones a Oxfam (Oxford Committee for Famine Relief) o del sufrimiento, su ausencia nos permite decir que no estamos frente a algo genuino.[4]

Un problema conexo es el inmenso poder de la autosugestión. Una vez que sabemos que un X debe ser un Y, afirmamos *y creemos* que es evidentemente un Y. Los expertos más grandes del mundo en Vermeer fueron embaucados por las (que *hoy* parecen ser) obvias falsificaciones de Van Meegeren. Proust se refiere a la «aptitud que permite adivinar las intenciones de un poema sinfónico, cuando se lee su explicación en el programa, y la cara de un niño cuando se conoce a sus padres». Un aficionado europeo al *jazz* modificó por completo su elevada estima de Jack Teagarden ni bien se enteró de que no era negro. Si estamos bien dispuestos hacia un escritor, quizá leamos significados profundos en lo que un lector imparcial consideraría comentarios triviales. Proyectamos nuestras expectativas en el mundo y luego sostenemos que éste confirma y justifica nuestras creencias.

La conclusión de estas observaciones es que debemos tener la precaución de no concebir las creencias, los deseos, las preferencias, las emociones y cosas por el estilo como entidades estables y duraderas en un pie de igualdad con las manzanas y los planetas. Los capítulos que siguen proporcionarán muchos ejemplos de la naturaleza elusiva, inestable o dependiente del contexto de los estados mentales. Dicho esto, los lectores también encontrarán enunciados que tal vez parezcan ejemplificar el tipo de seudoprecisión o rigor fingido contra el que he advertido. Para ir más allá de las declaraciones cualitativas que carecen de implicaciones definidas, como «los agentes atribuyen mayor peso al bienestar presente que al futuro», tenemos que decir algo sobre *la cantidad* de peso que atribuyen al presente. Una vez hecho esto, plantearemos de manera inevitable formulaciones que hacen distinciones más marcadas que las que observamos en el comportamiento de los agentes. El truco (más un arte que una ciencia) consiste en saber cuándo las simplificaciones iluminan y cuándo distorsionan.

Es menester hacer algunos comentarios finales sobre los estados y las operaciones mentales *inconscientes*. En este libro me referiré reiteradas veces al funcionamiento inconsciente de la mente. La reducción de la disonancia (Capítulo 1), el pensamiento desiderativo (Capítulo 7) y la trasmutación de los motivos (Capítulo 4), por ejemplo, son provocados por mecanismos inconscientes. Quizá no entendamos bien cómo funcionan, pero me parece imposible negar que existen. Muchos también han postulado la existencia de *estados* mentales inconscientes. El autoengaño,

como el pensamiento desiderativo, presupone que hay creencias inconscientes. En opinión de Freud, todos tenemos una serie de deseos inconscientes e inconfesables. Tal vez haya también emociones y prejuicios inconscientes.

En tanto y en cuanto los estados mentales inconscientes tienen eficacia causal, debería ser posible identificarlos por sus efectos. Si la fuerza con que se rechaza una afirmación está fuera de toda proporción, por ejemplo, podríamos inferir que la persona que la rechaza cree realmente en ella, aunque de manera inconsciente: «La señora protesta demasiado, a mi parecer». Hay una historia (que me ha sido imposible localizar) que se cuenta sobre Sigmund Freud, a quien invitaron a conocer al doctor X, persona de destacada actuación en el movimiento judío internacional. Durante la conversación, el doctor X le preguntó: «Dígame, doctor Freud, en su opinión, ¿cuál es la personalidad judía más importante en el mundo de nuestros días?» Freud respondió cortésmente: «Creo, desde luego, que debe de ser usted, doctor X». Cuando éste replicó: «No, no», Freud le preguntó: «¿Un solo "no" no habría bastado?» La doble negación puede ser equivalente a la afirmación.

También es posible identificar los prejuicios inconscientes por sus efectos. En unos experimentos, se instruyó ante todo a los sujetos para que distinguieran rápidamente (con una ligera palmada en la rodilla izquierda o derecha), entre los nombres de una lista, los que casi siempre se consideran negros (como Malik y Lashonda) y los que suelen verse como blancos (por ejemplo Tiffany y Peter). A continuación, se les solicitó que hicieran una rápida clasificación de una lista de palabras, distinguiendo las que tenían un significado agradable (como «amor» y «bebé») de las de significado desagradable (como «guerra» y «vómito»). Luego, los sujetos clasificaron una lista ordenada al azar que incluía los nombres negros, los nombres blancos, las palabras agradables y las palabras desagradables. En primer lugar se les pidió que se dieran una palmada en la rodilla izquierda para cualquier nombre negro o palabra de significado desagradable, y en la rodilla derecha para cualquier nombre blanco o palabra de significado agradable. Tras ello se modificaron las instrucciones. Se les pidió entonces que se golpearan la rodilla izquierda para los nombres blancos y las palabras desagradables y la rodilla derecha para los nombres negros y las palabras agradables. Los participantes tardaron casi el doble en cumplir la segunda tarea, aun cuando objetivamente la dificultad era igual en ambas.

Con frecuencia, las emociones inconscientes pueden ser identificadas por los observadores, que infieren su existencia de las expresiones fisiológicas o de comportamiento características. Casi todos hemos escuchado y muchos hemos proferido una airada sentencia: «No estoy enojado». La envidia puede manifestarse en un tono agudo y una tendencia a adoptar una postura despectiva que son obvios para los observadores, pero no para el sujeto. En *Rojo y negro*, Madame de Rênal sólo descubre sus sentimientos por Julien Sorel cuando sospecha que él puede estar enamorado de su doncella: de ese modo, una emoción (los celos) revela la presencia de otra (el amor).

En este aspecto, el autoengaño (véase el Capítulo 7) es más problemático. Supongamos que forjo y luego reprimo la creencia de que mi mujer tiene un amorío con mi mejor amigo. Aunque inconsciente, la creencia de que son amantes podría guiar mis acciones si, por ejemplo, me impidiera ir a la zona de la ciudad donde vive mi amigo para no correr el riesgo de ver a mi mujer en su compañía. Esto puede parecer una historia verosímil, pero por lo que sé no hay pruebas de que las creencias inconscientes tengan eficacia causal. Muchos argumentos que sostienen la existencia del autoengaño se basan en: 1) la exposición de la persona a pruebas que sugieren vigorosamente una creencia que ella no quiere que sea cierta, y 2) el hecho de que el agente profese y actúe en función de una creencia más aceptable. Para obtener pruebas directas de la persistencia inconsciente de la creencia inadmisible, sería preciso mostrar que 3) ésta también es capaz de guiar la acción, como en el ejemplo hipotético recién dado. Insisto, empero, en que no conozco ninguna demostración de este efecto.

Mi corazonada es que el fenómeno no existe. Es una placentera presunción imaginar que mis creencias inconscientes podrían ser servidoras de mis creencias conscientes, al apartarme de las pruebas capaces de socavarlas; no es ésta, empero, más que una historia del tipo «así es porque así es», sin base alguna. En una actitud igualmente especulativa, uno podría imaginar que «el inconsciente» es capaz de inducir estrategias indirectas (un paso atrás, dos pasos adelante), por ejemplo, y hacer que un niño se autoinflija una herida para llamar la atención de los padres. Estas sugerencias hacen que el inconsciente se parezca demasiado a la mente consciente, pues le atribuyen la capacidad de tener representaciones del futuro (Capítulo 6) y de las acciones e intenciones de otras personas

(Capítulo 19). Los estados mentales de los que no tenemos conocimiento pueden causar acciones espontáneas como contestar «no, no» en vez de un simple «no», pero no conozco prueba alguna de que también puedan generar un comportamiento instrumentalmente racional.

* * *

NOTA BIBLIOGRÁFICA

Amos Tversky y Daniel Kahneman, «Judgment under uncertainty: heuristics and biases», *Science*, 185, 1974, págs. 1124-1131, presentan pruebas del anclaje de las evaluaciones de probabilidad. Por su parte, Dan Ariely, George Loewenstein y Drazen Prelec, «"Coherent arbitrariness": stable demand curves without stable preferences», *Quarterly Journal of Economics*, 118, 2003, págs. 73-105, exhiben pruebas del anclaje de las preferencias. La referencia al obispo Andrewes procede de Adam Nicolson, *God's Secretaries: The Making of the King James Bible*, Nueva York, HarperCollins, 2003, y la de la sanación real proviene de Marc Bloch, *Les Rois thaumaturges: étude sur le caractère surnaturel attribué à la puissance royale particulièrement en France et en Angleterre*, París, Armand Colin, 1961 [*Los reyes taumaturgos*, México, Fondo de Cultura Económica, 1988]. Se encontrará un buen análisis del sentimentalismo en Michael Tanner, «Sentimentality», *Proceedings of the Aristotelian Society*, 77, número especial, 1976-1977, págs. 127-147. Brian Nosek, Mahzarin Banaji y Anthony G. Greenwald, «Harvesting implicit group attitudes and beliefs from a demonstration website», *Group Dynamics*, 6(1), 2002, págs. 101-115, informan de un experimento en gran escala (realizado en Internet) sobre los prejuicios inconscientes. Con respecto al argumento de que Freud forjó un inconsciente demasiado similar a la mente consciente, véase Lionel Naccache, *Le Nouvel inconscient: Freud, Christophe Colomb des neurosciences*, París, Odile Jacob, 2006.

Notas al pie

[1] Debemos suponer que el suceso en cuestión es tal que, de ocurrir, no afectará personalmente al individuo; por ejemplo, el descubrimiento de vida en otros planetas. Si el suceso es la victoria de su equipo deportivo favorito, el individuo tal vez apueste a su derrota, a fin de tener, pase lo que pasare, algo que lo complazca.

[2] Este enunciado es probablemente demasiado terminante. Mediante la manipulación del procedimiento se podría deducir cualquier atribución de probabilidad entre el 50% y el 80%, pero ninguna fuera de esa gama. En ese caso, sería lícito afirmar que el sujeto se inclina a creer que el suceso se producirá, pero no está seguro de que así sea. Sin embargo, esta evaluación es menos precisa de lo que requieren los modelos convencionales de toma de decisiones.

[3] La idea de la religión como «opio del pueblo» también sugiere que es un bien de consumo y no una premisa para la acción. Podría ser, sin embargo, una premisa para la inacción.

[4] Un fenómeno relacionado se produce cuando la gente adopta una perspectiva de tercera persona sobre sí misma. Un proverbio dice: «La virtud no se conoce a sí misma». Además, no podemos afirmar de manera coherente nuestra propia ingenuidad, dado que la idea misma presupone una falta de autoconciencia. Y como dice Nero Wolfe en una de las novelas de Rex Stout protagonizadas por este personaje: «Afirmar la dignidad es perderla».

Capítulo 4

Motivaciones

Este capítulo y los dos siguientes se dedicarán a variedades de la motivación. En éste, la discusión es bastante general. En los siguientes, me concentro en dos problemas específicos, el egoísmo contra el altruismo y la imprevisión temporal contra la previsión temporal. Estos dos problemas se complementan: el segundo, por decirlo así, es la versión intertemporal del primero, un contraste interpersonal. Como veremos, también están sustancialmente relacionados, en cuanto la previsión puede *remedar* el altruismo.

El conjunto de las motivaciones humanas es un pastel cuyas porciones pueden cortarse de muchísimas maneras. Aunque ninguna de ellas puede aspirar a un estatus canónico, hay cuatro enfoques que he considerado útiles. El primero propone un continuo de motivaciones; el segundo y el tercero muestran una tricotomía, y el cuarto, una dicotomía simple. Las clasificaciones, aunque similares, exhiben a la vez interesantes diferencias, y nos permiten iluminar el mismo comportamiento desde distintos ángulos.

De lo visceral a lo racional

El 11 de septiembre de 2001, algunas personas saltaron a la muerte desde el World Trade Center debido al calor abrumador. «A decir verdad, no habría que concebirlo como una elección», dijo Louis Garcia, jefe de bomberos de la ciudad de Nueva York. «Si uno pone a unas cuantas personas junto a una ventana e inyecta un calor semejante, es bastante

probable que la mayoría se sienta forzada a saltar.» No había una verdadera alternativa. En el plano subjetivo, ésta puede ser también la experiencia de quienes beben agua de mar cuando no tienen agua dulce a su alcance. Tal vez sepan que beber aunque sea un poco de agua de mar es el comienzo de un peligroso camino: cuanto más se bebe, más sed se siente. No obstante, para algunos la tentación quizá parezca irresistible. El ansia vehemente de sustancias adictivas también puede experimentarse de esta manera. Un escritor del siglo XVIII, Benjamin Rush, propuso una dramática ilustración: «Cuando uno de sus amigos lo instó vigorosamente a dejar la bebida, (un borracho consuetudinario) dijo: "Si hubiera un cuñete de ron en un rincón de la habitación, y entre él y yo un cañón que disparara sin cesar, no podría dejar de pasar frente al cañón para llegar hasta el ron"». El deseo sexual puede, igualmente, ser lo bastante abrumador para acallar inquietudes más prudenciales.

Es posible, asimismo, que ciertas emociones sean lo bastante fuertes para desplazar todas las demás consideraciones. El sentimiento de vergüenza, por ejemplo, puede ser insoportablemente doloroso, como lo mostró en 1996 el suicidio de un almirante de la marina estadounidense a quien estaban por desenmascarar por usar algunas medallas a las que no tenía derecho, o, en 1997, el de seis franceses que habían sido descubiertos como consumidores de material pedofílico. También la ira puede tener una fuerza abrumadora, como lo demostró Zinedine Zidane cuando, el 9 de julio de 2006, mientras se jugaban los últimos minutos de la final de la Copa del Mundo de fútbol, aplicó un cabezazo a un adversario italiano como represalia por una provocación, bajo la mirada de setenta mil personas en el estadio y un número estimado de mil millones de telespectadores en todo el planeta. Si se hubiera tomado un segundo para reflexionar, se habría dado cuenta que la acción podía costar la derrota de su equipo y el derrumbe de su reputación.

Excepto, quizás, en el caso del impulso por saltar del World Trade Center, es dudoso que alguno de estos deseos fuera literalmente irresistible, a la manera en que una gran piedra que rueda cuesta abajo podría serlo para una persona que tratara de interrumpir su carrera. (La urgencia de dormirse puede ser irresistible, pero el quedarse dormido no es una acción; por eso los intentos de conseguirlo son contraproducentes.) Los adictos son un tanto sensibles a los costes: consumen menos cuando los precios suben.[5] Las personas a bordo de un bote salvavidas pueden a veces impedirse unas a

otras tomar agua de mar. La tentación sexual y el impulso de matarse por vergüenza son sin duda resistibles. Debido a su intensidad, estos anhelos viscerales se sitúan, no obstante, en un extremo del espectro de las motivaciones humanas. Tienen el potencial, no siempre realizado, de bloquear la deliberación, las soluciones de compromiso y hasta las decisiones.

En el otro extremo, tenemos un paradigma o, más bien una caricatura, del agente racional a quien no perturban los factores viscerales, incluida la emoción. Este agente actúa sólo después de haber sopesado con mucho cuidado, pero no más del que justifican las circunstancias, las consecuencias de cada alternativa disponible en comparación con las demás. Un general, un presidente de empresa o un médico racionales se ocupan meramente de encontrar la mejor manera de alcanzar una meta objetiva como ganar la guerra, maximizar las ganancias o salvar una vida. Los deseos subjetivos y sus raíces viscerales no entran en la ecuación. Aunque en el Capítulo 11 veremos que el concepto de racionalidad es mucho más amplio que esta idea exangüe, ésta quizá sirva como punto de referencia.

La diferencia entre el *miedo* visceral y el *miedo* prudencial es un ejemplo de la distinción entre la motivación visceral y la motivación racional. Si bien es corriente referirse al miedo como una emoción, tal vez sólo sea un complejo de creencias y deseos. Cuando digo: «Me temo que va a llover», quiero decir *únicamente* que creo que va a llover y que desearía que no lloviera. Si el «temor» inspira la acción, como ocurre cuando tomo un paraguas para protegerme de la lluvia, es un paradigma del comportamiento racional (Capítulo 11). Ninguno de los rasgos característicos de las emociones (Capítulo 8) está presente. El miedo visceral, en contraste, puede inducir una acción que no sea instrumentalmente racional. Se ha calculado, por ejemplo, que trescientos cincuenta estadounidenses que de lo contrario no habrían muerto, perdieron la vida en las carreteras por evitar el riesgo de volar luego del 11 de septiembre de 2001. En cambio, no parece que los españoles hayan incurrido en un número excesivo de muertes por cambiar el ferrocarril por el automóvil después de los ataques a los trenes ocurridos en Madrid el 11 de marzo de 2004. Es posible que, luego de la extensa cadena de ataques de ETA, la población haya desarrollado una actitud de temor prudencial y no de miedo visceral hacia las bombas del terrorismo. Para los españoles, los

ataques terroristas pueden haber sido un riesgo más entre otros, similar al de la lluvia, aunque más peligroso.

Entre los extremos del continuo de lo visceral a lo racional, encontramos un comportamiento que es motivado en parte por factores viscerales, pero también es un tanto sensible a las consideraciones de costes y beneficios. Un hombre puede buscar venganza (un deseo visceral) y, no obstante, esperar también su oportunidad hasta estar en condiciones de coger desprevenido a su enemigo (una inquietud prudencial). Si reta a duelo a éste (tal cual lo exigen las normas del honor), tal vez tome lecciones de esgrima en secreto (una práctica poco honorable pero útil). Si una persona recibe una oferta que es a la vez injusta y ventajosa, en el sentido de que estará mejor si la acepta, puede aceptarla o rechazarla, según cuál sea la fuerza de su interés en comparación con la fuerza de su resentimiento.[6] En casos más complejos, un factor visceral podría contrarrestar otro. El deseo de tener una aventura sexual extramatrimonial podría ser neutralizado por los sentimientos de culpa. Un impulso de lucha, provocado por la ira, puede compensar o anticiparse a un impulso de huida, generado por el miedo.

Interés, razón y pasión

En su análisis de las motivaciones humanas, los moralistas franceses del siglo XVII hacían una fructífera distinción entre interés, razón y pasión. El interés es la búsqueda de la ventaja personal, trátese de dinero, fama, poder o salvación. Aun los actos para ayudar a nuestros hijos se consideran como una promoción del interés, por la íntima relación existente entre nuestro destino y el suyo. Un padre que envía a sus hijos a una costosa escuela privada donde éstos pueden recibir la mejor educación no sacrifica su interés: lo promueve. Entre las pasiones pueden estimarse incluidas las emociones, al igual que otros impulsos viscerales como el hambre, la sed y las ansias sexuales y adictivas. Los antiguos también incorporaban los estados de locura a la misma categoría general porque, como las emociones, son involuntarios y espontáneos y subvierten la deliberación racional. A muchos efectos, es posible incluir asimismo entre las pasiones los estados de intoxicación. Desde el punto de vista del derecho, la ira, la ebriedad y la locura han sido con frecuencia tratadas en un pie de igualdad.

La razón es una idea más complicada. En su mayor parte, los moralistas la utilizaban (como yo lo haré aquí) en relación con el deseo de promover el

bien público en vez de los fines privados. De vez en cuando, también la usaban para referirse a motivaciones (prudenciales) de largo plazo, distinguidas de las inquietudes (miopes) de corto plazo. Ambas ideas pueden resumirse bajo el encabezado de la *imparcialidad*. Al elaborar las políticas públicas, habría que tratar a las personas de manera imparcial en vez de favorecer a algunos grupos o individuos en detrimento de otros. También los individuos pueden actuar sobre la base de esa motivación. Los padres quizá sacrifiquen su interés al enviar a sus hijos a una escuela pública, porque creen en la igualdad de oportunidades. Al mismo tiempo, los planificadores de políticas, así como los individuos particulares, deberían tratar en forma imparcial los resultados que se producen en ocasiones sucesivas, atribuyendo a cada uno el mismo peso en las decisiones actuales, en vez de privilegiar los resultados previstos para el futuro cercano. De hecho, según argumentaron algunos moralistas, una preocupación por el interés de largo plazo también tenderá a promover el bien público. En la Convención Federal celebrada en Filadelfia, por ejemplo, George Mason sostuvo que

> es nuestro deber prestar atención a los derechos de toda clase de personas. Quien habla se ha extrañado con frecuencia ante la indiferencia de las clases superiores de la sociedad frente a este dictado de humanidad y política; considerando que, por opulentas que fueran sus circunstancias o elevada su situación, el transcurso de algunos años no sólo podría distribuir, sino que indudablemente distribuiría su posteridad a través de las clases más bajas de la sociedad. En consecuencia, todo motivo egoísta, todo lazo familiar, deberían sugerir un sistema político que velara no menos cuidadosamente por los derechos y la felicidad de los órdenes más bajos de la sociedad que por los de sus órdenes más altos.

Una y otra forma de imparcialidad tienen grados. La fuerza de la preocupación por los otros tiende a variar inversamente no sólo con la distancia genealógica, sino con la lejanía geográfica. De manera similar, incluso los individuos prudentes suelen atribuir un poco más de peso al futuro cercano que al más remoto, un hecho que se explica sólo en parte por su conciencia de que tal vez no vivan para disfrutar de ese tiempo distante.

Como ejemplo del modo de entender el comportamiento en términos de cualquiera de estas tres motivaciones, podemos citar una carta de 1783 de Robert Livingston, juez canciller de Nueva York, a Alexander Hamilton, en la cual aquél comenta la persecución emprendida contra quienes habían tomado partido por los británicos durante las guerras de la independencia:

> Lamento seriamente con usted el violento espíritu de persecución que predomina aquí y temo sus consecuencias sobre la riqueza, el comercio y la tranquilidad futura del estado. Y tanto más me

> duele cuanto que, a mi juicio, apenas intervienen en él *motivos patrióticos más puros*. En algunos pocos es un ciego espíritu de *venganza y resentimiento*, pero en la mayoría es el más *sórdido interés*.

Las palabras que he puesto en bastardilla corresponden a la razón, la emoción y el interés, respectivamente. Los adjetivos son reveladores: la razón es pura, la pasión es ciega y el interés es sórdido. Volveré a algunas implicaciones de estas valoraciones.

Ello, yo, superyó

En sus análisis de las motivaciones humanas, Freud también distinguió tres formas básicas, cada una de ellas vinculada a un subsistema diferente de la mente. Los tres sistemas son el ello, el yo y el superyó, correspondientes al principio de placer, el principio de realidad y la conciencia, respectivamente. El ello y el superyó representan los impulsos (el ello) y su control (el superyó), mientras que el yo, «desvalido hacia ambos costados […], se defiende en vano de las insinuaciones del ello asesino y de los reproches de la conciencia moral castigadora». En un enunciado más esclarecedor del mismo artículo («El yo y el ello»), Freud señala que el yo es «una pobre cosa sometida a tres servidumbres y que, en consecuencia, sufre las amenazas de tres clases de peligros: de parte del mundo exterior, de la libido del ello y de la severidad del superyó». Sin embargo, ni siquiera esta formulación expresa del todo lo que es, a mi entender, el núcleo útil de la idea freudiana. Me refiero a la proposición de que, mientras navega por el mundo externo (el principio de realidad), el yo también tiene que librar una guerra de dos frentes contra los impulsos del ello (principio de placer) y el severo control punitivo de los impulsos ejercido por el superyó (conciencia).[7]

Esta proposición es original, profunda y verdadera. Lo que le falta es un mecanismo. ¿Por qué no podría el mismo yo ejercer el control de los impulsos que fuera necesario? ¿Por qué la moral y la conciencia adoptan tantas veces la forma de reglas rígidas? ¿Es preciso estipular la existencia de funciones mentales separadas y casi autónomas? Fue necesario el trabajo pionero de George Ainslie para brindar una respuesta satisfactoria a estas preguntas. Examino sus concepciones en el Capítulo 13. Aquí, quiero limitarme a destacar el hecho de que muchos impulsos deben mantenerse a raya debido al daño *acumulativo* que pueden infligir si no se los controla.[8]

En cualquier oportunidad dada, el exceso de bebida o comida, el derroche de dinero o la dilación (la omisión de las tareas escolares, por ejemplo) no tienen por qué ser muy dañinos para el agente. El daño se produce luego de reiterados excesos (o reiteradas omisiones). En consecuencia, el eje del control de los impulsos no debe estar en la ocasión individual, dado que la persona siempre puede decirse que mañana comenzará una nueva vida mejor. Ese control debe afrontar el hecho de que el impulso surgirá previsiblemente en una cantidad indefinida de ocasiones. La solución está en la reformulación del problema, de modo tal que la omisión del control de un impulso en cualquier oportunidad concreta se tome como un predictor de esa misma omisión en todas las oportunidades posteriores. «Sí, puedo posponer el control de los impulsos hasta mañana sin incurrir en un perjuicio o un riesgo importantes, pero ¿por qué mañana va a ser diferente de hoy? Si fallo hoy, también fallaré mañana.» Al instaurar un *efecto interno de dominó* y elevar así las apuestas, el agente puede ganar una motivación para controlar sus impulsos de la que carecería si se limitara a tomar un día por vez. La otra cara de la moneda es que el control debe ser inflexible y, como decían los moralistas victorianos, «no tolerar una sola excepción».

La consideración de las consecuencias

Por último, las motivaciones pueden ser consecuencialistas o no consecuencialistas, es decir, orientadas o bien hacia el resultado de la acción o hacia la acción misma. Gran parte del comportamiento económico es puramente consecuencialista. Cuando la gente ahorra dinero para su vejez o los corredores bursátiles compran y venden acciones, no atribuyen un valor intrínseco, positivo o negativo, a dichos actos; sólo les interesan los resultados. En contraste, el pacifista incondicional que se niega a hacer el servicio militar aun contra el más malvado de los enemigos no toma en cuenta las consecuencias de su comportamiento. Para él, lo que cuenta es que algunas acciones están incondicionalmente prohibidas, por ejemplo quitar la vida a un ser humano. No se trata de que *desconozca* las consecuencias, como puede suceder en el caso de la acción emocional; se trata de que esas consecuencias no modifican su manera de actuar.

Las políticas públicas también pueden adoptarse sobre la base de uno u otro tipo de motivación. Un planificador de políticas podría hacer suyo el

principio de que «quien lo encuentra se lo queda» (por ejemplo, en las leyes de patentes), con el supuesto de que, si se otorga a la persona que descubre un nuevo recurso valioso la propiedad de éste, se descubrirán más recursos valiosos. Éste es un argumento de tipo consecuencialista. Un argumento no consecuencialista para la misma política podría sostener que la persona que descubre un nuevo recurso, ya se trate de un territorio o de una cura para el cáncer, tiene un *derecho* natural de propiedad con respecto a él. Para mencionar otro ejemplo, podemos considerar el discurso (XXXI) de Dión Crisóstomo contra la práctica de los habitantes de Rodas de reutilizar viejas estatuas de bronce para honrar a los benefactores de la ciudad: esa actitud, sostenía Dión, significaba a la vez violar los derechos de aquellos en cuyo honor las estatuas habían sido originalmente erigidas y desalentar a potenciales nuevos benefactores, que sabrían que las estatuas levantadas en su honor no tardarían en reciclarse en favor de algún otro. Es posible que los argumentos consecuencialistas justifiquen (o parezcan justificar) las medidas severas contra los terroristas, aun cuando los pasos que se den violen los valores no consecuencialistas asociados a los derechos humanos y las libertades civiles.[9]

Un caso especial de motivación no consecuencialista es el principio al que me referiré bajo diversos nombres, como kantismo cotidiano, imperativo categórico o pensamiento mágico: *haz lo que sería óptimo si todos hicieran lo mismo*. En cierto sentido, este principio está ligado a las consecuencias, puesto que el agente hace lo que provocaría el mejor resultado si todos los demás hicieran otro tanto. Esas consecuencias, sin embargo, no son las de *su* acción, sino las de un conjunto hipotético de acciones llevadas a cabo por él y por otros. En un caso determinado, la actuación sobre la base del principio podría tener consecuencias desastrosas para todos si los otros *no* hicieran lo mismo. En la arena internacional, el desarme unilateral es un ejemplo.

Otro caso es el siguiente principio de la ética judía. Supongamos que el enemigo está a las puertas y dice: «Dame a uno de los tuyos para matarlo y exceptuaremos a todos los demás; si te niegas, los mataremos a todos». El Talmud exige que en tales casos todos los judíos se dejen matar en vez de designar a uno para morir a fin de que el resto se salve. Sin embargo, si el enemigo dice «dame a fulano» bajo la misma condición, es aceptable entregarlo. Lo que está prohibido no es hacer que una persona muera para

salvar a otras, sino seleccionar a la víctima. La novela *La decisión de Sophie* presenta el mismo dilema.

Las normas sociales (Capítulo 22) representan otro caso especial de comportamiento no consecuencialista, con una peculiaridad importante. Dichas normas dicen a la gente qué hacer, por ejemplo vengarse por un insulto o abstenerse de comer un cabrito hervido en la leche de su madre, no porque sea posible obtener resultados deseables, sino porque la acción es imperativa en sí misma.[10] Si bien no se emprenden para *provocar* ningún resultado, puede considerarse que esas acciones se llevan a cabo para *impedir* un resultado, a saber, que otros nos censuren por no haberlas realizado. Cabe preguntarse entonces, con todo, si las censuras también se plantean por razones consecuencialistas similares. En general, yo diría que no. Por otra parte, cuando las acciones de otros dañan a una persona, ésta toma represalias aun en interacciones únicas en condiciones de completo anonimato, como las que pueden existir en ámbitos experimentales. Como la interacción carece de réplica, la persona no tiene nada que ganar en encuentros ulteriores, y como es anónima, no tiene por qué temer la censura de un tercero. Volveré a ocuparme de estos experimentos en varios otros capítulos.

Aun para un no consecuencialista declarado, las consecuencias pueden significar algo si son suficientemente importantes. Tómese un principio que muchos considerarían incondicional, la prohibición de torturar niños. En un escenario de «bomba de tiempo», imaginemos que una condición necesaria y suficiente para impedir la detonación de un artefacto nuclear en el centro de Manhattan es torturar al hijo pequeño de una terrorista en su presencia. *Si* la perspectiva pudiese ser verosímil, muchos no consecuencialistas tal vez admitirían la tortura. Otros dirían que, como las condiciones de ese escenario nunca existirán en la práctica, la prohibición absoluta sigue en vigor. Y otros proscribirían la tortura aun cuando el escenario se convirtiese en realidad. Mi tarea no consiste aquí en defender una de estas conclusiones, sino en hacer la observación empírica de que, en las situaciones de la vida real, lo que está en juego pocas veces es tan grave como para forzar al no consecuencialista a considerar las consecuencias de su comportamiento. Es posible que, de haber más cosas en juego, abandone sus principios, pero puesto que la situación no se plantea, no podemos decir con certeza si estamos simplemente ante una solución de compromiso

exorbitante o ante una total negativa a admitir las soluciones de compromiso.

Estos cuatro enfoques de la motivación dan cabida a algunos de los mismos fenómenos. Los factores viscerales, las pasiones y el principio de placer tienen a todas luces mucho en común. El último es válido para una gama más amplia de casos, porque además de la búsqueda del placer implica la evitación del dolor. Cuando los estudiantes dilatan el momento de hacer sus tareas escolares, no es necesariamente porque se sientan apasionados por hacer otra cosa. Con frecuencia, no hacen sino tomar el camino del menor esfuerzo. El superyó, la razón y las motivaciones no consecuencialistas también tienen algunas características en común. Aunque no todos los sistemas morales son rígidos e inflexibles, hay algunos que sí lo son. La teoría moral de Kant es un ejemplo notorio. (De hecho, es posible que su filosofía moral haya tenido su origen en las reglas privadas que elaboró para sí mismo con el fin de controlar sus impulsos; un ejemplo es la máxima de no fumar nunca más de una pipa después del desayuno.) Al mismo tiempo, la moral puede elevarse por encima de la rigidez, en los individuos no sometidos a la aversión de la ambigüedad. En realidad, a menudo se dice que la tolerancia de la ambigüedad es la marca distintiva de un yo saludable. En contraste, la relación entre la racionalidad, el interés, el yo y el consecuencialismo es más tenue. Sería absurdo pretender que la marca distintiva de un yo saludable es la búsqueda racional del egoísmo.

Querer y desear

Muchas veces concebimos que las motivaciones toman la forma del *querer generar* algún estado de cosas. Sin embargo, también pueden adoptar la forma del *desear la existencia de algún estado de cosas*. Esta distinción entre el querer y el deseo es importante si examinamos el componente motivacional de la emoción (Capítulo 8). Las emociones, de hecho, pueden estar acompañadas sea de un querer hacer algo, sea de un deseo de que algo suceda. En la ira o la *furia*, el impulso de A de vengarse de B no puede satisfacerse por el hecho de que C haga a B lo que A había planeado hacer, o por el hecho de que B sufra un accidente. Lo que importa no es simplemente el resultado, que B sufra, sino que sufra en virtud de la agencia de A. En el sadismo, lo que importa es también hacer sufrir al otro, no meramente que sufra. En contraste, lo que cuenta en el *odio* es que la

persona o grupo odiado desaparezca de la faz de la Tierra, ya sea por mi agencia o por la de algún otro. En la *malicia*, importa que el otro sufra, no que yo lo haga sufrir. De hecho, una persona maliciosa puede recular antes de ponerse en actividad y tomar medidas para hacer sufrir al otro, no sólo porque teme ser vista haciéndolo, sino porque esa actitud sería incompatible con la imagen que tiene de sí misma. Esta situación es aún más clara en la *envidia*. Muchas personas que disfrutarían al ver que un rival pierde sus posesiones y no harían nada para impedirlo aunque pudieran, jamás tomarían medidas concretas para destruirlo, aun cuando pudiesen hacerlo sin costes ni riesgos para sí mismas.[11] Una persona que no incendiaría la casa de su vecino quizá se abstuviera de llamar a los bomberos si la viera ardiendo.

El pensamiento desiderativo (Capítulo 7) se basa más en deseos que en voluntades. En algunos casos, el agente evita el duro trabajo de conformar el mundo a sus deseos y toma, en cambio, el fácil camino de adoptar una creencia apropiada sobre el mundo. Si deseo obtener un ascenso pero soy renuente a hacer un esfuerzo, tal vez me apoye en signos insignificantes para persuadirme de la inminencia de una mejora laboral. En otros casos, no existe la opción de actuar sobre el mundo. Tal vez yo sea incapaz de lograr que mi amor sea correspondido o que mi hijo enfermo se recupere. En tales casos, puedo o bien forjarme fantasías gratificantes o enfrentar los hechos. Una distinción más es la que se traza entre los casos en que las fantasías no tienen otras consecuencias para la acción y aquellos en que se las utiliza como premisas del comportamiento. Quizá me induzca a creer que una mujer que conozco abriga una pasión secreta por mí, pese a lo cual no le haré ninguna insinuación, sea porque la moral (o el egoísmo) me refrena, sea porque aliento la creencia ilusoria principalmente por su valor de consumo.[12] La ilusión también puede expresarse francamente a su objeto, como sucedió cuando una secretaria de John Maynard Keynes le dijo a éste que no podía evitar darse cuenta de su gran pasión por ella. Y así se arruinó la vida.

Estados que son esencialmente subproductos

Un factor que complica la distinción entre el deseo y el querer es que en algunos casos puedo conseguir X haciendo A, pero sólo si hago A a fin de conseguir Y. Si trabajo con empeño para explicar la base neurofisiológica

de la emoción y lo logro, tal vez me granjee una elevada reputación. Si me lanzo a trabajar por una causa política, quizá descubra al final del proceso que también he ganado un «carácter». Si toco bien el piano, puedo impresionar a otros. Estos beneficios indirectos son parasitarios de la meta principal de la actividad. Si mi motivación como académico es granjearme una reputación, es menos probable que lo consiga. La militancia en un movimiento político *únicamente* por los efectos de toma de conciencia o construcción del carácter sobre uno mismo está condenada al fracaso, o sólo será un éxito por azar. Como señaló Proust, un músico «puede a veces traicionar [su verdadera vocación] en bien de la gloria, pero al buscar así la gloria, se aparta de ella y sólo la encuentra cuando le da la espalda». La autoconciencia interfiere en el desempeño. Como también escribió Proust, aunque «la mejor manera de hacer que nos busquen es ser difíciles de encontrar», nunca dio a nadie un consejo en ese sentido, pues «este método de alcanzar el éxito social sólo funciona si uno no lo adopta con ese fin».

La gloria musical o el éxito social se incluyen en la categoría de los *estados que son esencialmente subproductos*: estados que no pueden realizarse a través de acciones sólo motivadas por el deseo de realizarlos. Se trata de estados que pueden *suceder*, pero no *provocarse* intencionalmente mediante una mera decisión. Entre ellos se cuentan el deseo de olvidar, el deseo de creer, el deseo de desear (por ejemplo, el de superar la impotencia sexual), el deseo de dormir, el deseo de reír (uno no puede hacerse cosquillas) y el deseo de vencer el tartamudeo. Es probable que los intentos de realizar los deseos sean ineficaces y puedan incluso empeorar las cosas. Entre los moralistas y los novelistas es un lugar común decir que el hedonismo intencional es contraproducente,[13] y que nada graba una experiencia en la memoria tan profundamente como el intento de olvidarla. Aunque *deseemos* que estos estados se realicen, debemos estar en guardia contra el *querer* realizarlos.

Mucha gente se preocupa por la *salvación* (en la vida después de la muerte) y la *redención* (por los males que ha cometido). Quizá también crea que puede alcanzar esos objetivos mediante la acción. Sufrir la muerte de un mártir en la lucha contra los infieles puede representar el pasaporte al cielo; al menos, así lo creen algunos. Combatir contra los nazis luego de haber colaborado con ellos en una etapa anterior puede redimir de la maldad. No obstante, si esas acciones se emprenden con el *propósito* de alcanzar la salvación o la redención, es posible que fracasen. En la teología

católica, la intención de comprar un lugar en el cielo por medio del martirio voluntario sería un ejemplo del pecado de simonía. Algunos eruditos islámicos plantean una crítica similar con respecto a los atacantes suicidas motivados por la creencia de que conseguirán un lugar privilegiado en el Paraíso. Montaigne escribe que cuando los espartanos

> hubieron de decidir a qué individuo correspondía el honor de haber luchado mejor aquel día, consideraron que Aristodemo se había arriesgado más valerosamente que ninguno; sin embargo, no le premiaron por ello porque su valor había sido incitado por un deseo de desquitarse de los reproches de los que se había hecho merecedor en la batalla de las Termópilas.

Jean Prouvost, el magnate francés de la prensa que había colaborado con las fuerzas alemanas durante la ocupación de Francia, trató de redimirse firmando un cuantioso cheque para la Resistencia cuando resultó claro que los alemanes estaban perdiendo la guerra. Tras la liberación, el Tribunal Superior le otorgó un *non-lieu* (una sentencia que suspende, anula o retira un caso sin llevarlo a juicio), algo que, es de presumir, los espartanos no habrían hecho.[14]

Expulsión contra atracción

¿Por qué la gente deja un país por otro? ¿Por qué los académicos cambian de universidad? A menudo, las respuestas se clasifican bajo la fórmula de «expulsión contra atracción». Uno puede emigrar porque la situación en su país es intolerable o porque la situación en el extranjero es irresistiblemente seductora; ésta es al menos una forma corriente de ver el asunto. En muchas situaciones, sin embargo, se trata de una forma engañosa. Por lo común, la gente emigra porque *compara* la situación en su país y en el exterior y comprueba que la diferencia es lo bastante grande para justificar una mudanza, aun cuando se tomen en cuenta los costes de ésta.[15] Sin embargo, puede tener sentido hacer una distinción entre los motivos de expulsión y los motivos de atracción, cuando los primeros están más cerca del extremo visceral del continuo y los segundos, más cerca del extremo racional. Las personas embargadas por un miedo intenso a veces escapan del peligro en vez de correr hacia la seguridad. Sólo piensan en huir, y no se detienen a considerar si no salen de Guatemala para entrar en Guatepeor. Dependientes de la droga, los adictos pueden estar motivados por la atracción de la euforia (cocaína) o la expulsión de la disforia

(heroína). También el comportamiento suicida puede deber más a la expulsión que a la atracción. Es un escape de la desesperación, no una fuga hacia algo.

El funcionamiento de las normas sociales (Capítulo 22) puede verse asimismo en función del contraste entre expulsión y atracción. El deseo de sobresalir de una manera socialmente aprobada ejerce una fuerte atracción en muchos individuos, ya se afanen por la *gloria* (ser los mejores) o por el *honor* (triunfar en una competencia o un combate). A otros individuos les interesa más evitar la vergüenza asociada a la violación de las normas sociales. En algunas sociedades hay una norma general que dice: «No asomes la cabeza». Sobresalir en algo es desviarse, y la desviación es objeto de una desaprobación universal: «¿Por quién se toma?» La fuerza relativa de estas dos motivaciones varía a través de las sociedades y dentro de cada una de ellas. La Atenas clásica ilustra el afán competitivo por la excelencia.[16] En las sociedades modernas, las ciudades pequeñas exhiben con frecuencia los efectos asfixiantes de la hostilidad a la excelencia. Para aventurar una generalización, diremos que en general el carácter expulsivo de la vergüenza parece ser una motivación más importante que el carácter atractivo de la gloria, lo cual no significa decir que este último no pueda ser poderoso.

Conflicto motivacional

La existencia de motivaciones *antagónicas* es un lugar común:

Tengo una necesidad tan imperiosa de un libro que siento la tentación de hurtarlo de la biblioteca, pero también quiero comportarme moralmente.

Frente a un pendenciero, siento a la vez temor y enojo: quiero correr, pero también darle un golpe.

Quiero que todos los niños tengan educación pública, pero también quiero que mi hijo vaya a una escuela privada para tener la mejor educación.

Quiero un candidato que esté a favor del aborto legal, pero también quiero uno que propicie la rebaja de los impuestos.

Quiero fumar, pero también estar sano.

Si me hacen un ofrecimiento ventajoso pero turbio, en términos de «tómalo o déjalo», quiero rechazarlo porque es turbio pero, a la vez, aceptarlo porque es ventajoso.

Quiero hacer una donación a una institución de beneficencia, pero también promover mis propios intereses.

Estoy tentado de tener una aventura extramatrimonial, pero también quiero preservar mi matrimonio.

¿Cómo se resuelve el conflicto entre estas motivaciones? Una respuesta general podría afrimar lo siguiente. Cuando la situación es tal que «el ganador toma todo», de modo que no hay compromiso (material) posible, se impone la motivación más fuerte.[17] Si mi preocupación por mi hijo es más fuerte que mi preocupación por la escolarización de los niños en general, enviaré a aquél a una escuela privada. Si mi inquietud favorable a la libre elección en materia de aborto es más fuerte que mi interés en el recorte de impuestos y ningún candidato propicia ambas posiciones, votaré por un candidato abortista que proponga elevar la presión impositiva. Si alguien me ofrece tres dólares de un pozo común de diez dólares, y pretende quedarse con el resto, los acepto. Si me ofrecen sólo dos dólares, rechazo el ofrecimiento.[18] Cuando un compromiso es posible, la motivación más fuerte tiene un impacto más intenso que la menos fuerte. Un fumador quizá reduzca su consumo de cigarrillos de treinta a diez por día. Como reflejo de la fortaleza de mi altruismo, tal vez yo destine el 5% de mis ingresos a la beneficencia.[19]

Esta respuesta no es exactamente errónea, pero sí bastante simplista, toda vez que la idea de «fuerza de la motivación» es más complicada de lo que dan a entender estos rápidos ejemplos. Una motivación puede deber su fortaleza a su mera intensidad psíquica; en ese sentido, los motivos viscerales son a menudo más fuertes que «la moderada voz de la razón», según las palabras de Madison. Sin embargo, una motivación fuerte también puede conquistar la vigorosa adhesión del agente debido al elevado valor que tiene en su sociedad. De hecho, cada sociedad o cultura se caracteriza por una jerarquía normativa de motivaciones. En igualdad de las demás condiciones, una persona llevará a cabo una acción dada por el motivo A y no por el motivo B si A ocupa un lugar más alto en la jerarquía. Éstas son *metamotivaciones*, deseos de ser animado por deseos de cierto tipo.[20] Aunque más débiles en el sentido visceral, tal vez se impongan a la larga sobre otras motivaciones.

El interés y la pasión, sobre todo, muestran a menudo cierta *deferencia a la razón*.[21] Como dijo Séneca: «La razón desea que la decisión que toma sea justa; la ira desea que parezca justa la decisión que ha tomado». Puesto que hay muchas concepciones de la razón, de la justicia y de la equidad que parecen plausibles, con frecuencia será efectivamente posible presentar como conforme a la razón una decisión tomada en un arranque de ira. Los juicios a los colaboracionistas en los países que Alemania había ocupado

durante la Segunda Guerra Mundial se fundaron en muchos casos en un profundo deseo de venganza. No obstante, debido a su deferencia a la razón, combinada con su deseo de diferenciarse de las prácticas ilegales de los regímenes ocupantes, los nuevos líderes presentaron las severas medidas como si se basaran en la justicia y no en la emoción. Una persona puede tener un interés de primer orden en no hacer donaciones a la beneficencia y un deseo de segundo orden de no verse como si sólo la moviera el interés. En deferencia a la razón, quizás haga suya entonces la filosofía de la beneficencia (Capítulo 2) que puede justificar pequeñas donaciones. Si otros dan mucho, la persona en cuestión adoptará un enfoque utilitario que justifique las pequeñas donaciones, y si otros dan poco, adoptará un enfoque basado en la equidad que justifique el mismo comportamiento.

En estos casos, la razón no tiene un papel causal independiente. Sólo induce una justificación a posteriori de acciones ya decididas con otros fundamentos. El conflicto no se resuelve: no se hace más que barrerlo bajo la alfombra. En otros casos, la búsqueda de una justificación basada en la razón puede modificar el comportamiento. Si me inclino por un enfoque de la beneficencia basado en la equidad porque otros dan poco, y éstos comienzan de improviso a donar mucho más generosamente que yo, tengo que seguir su ejemplo. *La misma necesidad de autoestima* que en un primer momento me ha llevado a justificar un comportamiento egoísta mediante consideraciones imparciales también me impide modificar mi concepción de la imparcialidad cuando ya no actúa en mi favor. Es posible imaginar que, en *Rey Lear*, el rey de Borgoña y el rey de Francia se enamoran inicialmente de Cordelia por las perspectivas que implica, pero que sólo el primero se preocupa tan poco por su autoestima que puede dar rienda suelta a la emoción cuando ya no coincide con su interés. Éste es un caso en que el interés rinde deferencia a la pasión y no a la razón, y sugiere que la primera o, mejor, esa pasión en particular, se sitúa por encima del interés en la jerarquía normativa. Otras pasiones, como la envidia, bien podrían estar por debajo del interés. Tal vez constataríamos entonces esfuerzos por emprender sólo la acción basada en la envidia que pueda presentarse de manera verosímil como basada en el interés. Las acciones que no puedan verse bajo esa luz no se llevarán a cabo.

La teoría de la disonancia cognitiva predice que cuando una motivación es *apenas un poco* más fuerte que otra, procurará reclutar aliados con el fin de que las razones de un lado resulten decisivamente más fuertes. La mente

inconsciente sale de compras, por así decirlo, en busca de argumentos adicionales en favor de la conclusión tentativa alcanzada por la mente consciente.[22] En tales casos, la «fuerza» de la motivación no puede darse por descontada y debe vérsela en cambio, al menos hasta cierto punto, como un producto del proceso mismo de toma de decisiones. Supongamos que, al comprar un auto, asigno diferentes valores a distintas características (velocidad, precio, confort, apariencia) de cada una de las alternativas y llego a una evaluación global mediante la comparación de las sumas ponderadas de dichos valores. Podría, por ejemplo, asignar un valor global de cincuenta a la marca A y de cuarenta y ocho a la marca B. Debido a la incómoda paridad de la comparación, modifico inconscientemente los valores a fin de que A gane con claridad, con sesenta contra cuarenta y cinco como resultado global. Antes de concretar la compra, doy con la marca C, que con los viejos valores habría obtenido cincuenta y cinco, pero que con los nuevos sólo llega a cincuenta. Si yo hubiera visto las alternativas en el orden C-A-B, habría elegido C. Como las he encontrado en el orden A-B-C, escojo A. Esa *dependencia del camino elegido* socava la simple idea de que los conflictos motivacionales se resuelven de conformidad con la fuerza motivacional dada.

Según lo que he calificado de concepción simplista, la decisión de hurtar o no un libro de la biblioteca podría representarse de la siguiente manera. En un platillo de la balanza está el beneficio de poder utilizar el libro; en el otro, el coste de los sentimientos de culpa. Lo que termine por hacer dependerá exclusivamente de que el coste exceda el beneficio o a la inversa. Pero esto no puede estar bien; supongamos, en efecto, que alguien me ofrece una «píldora de la culpa» que suprime todos los sentimientos aflictivos y culpables por hurtar el libro. Si la culpa interviniera en mis decisiones como un mero coste psíquico, sería racional tomar la píldora, así como lo sería tomar una píldora que me evitara una resaca luego de una parranda alcohólica programada. Sostengo, sin embargo, que la mayoría de la gente *sentiría tanta culpa por tomar la píldora como por hurtar el libro*.[23] No niego que pueda haber, en cierto sentido, una solución de compromiso entre la moral y el egoísmo; sólo digo que no es posible representarla de esa manera simplista.

Ahora, un ejemplo más complejo. Ojalá yo deseara no desear no comer torta de crema. Quiero comer torta de crema porque me gusta. Desearía que no me gustara, porque, como cualquier persona moderadamente vanidosa,

creo que es más importante mantenerse delgado. Pero desearía ser menos vanidoso. *Sin embargo, ¿ese deseo se activa únicamente cuando quiero comer torta de crema*? En el conflicto entre mi deseo de torta de crema, mi deseo de ser delgado y mi deseo de no ser vanidoso, el primero y el último pueden formar una alianza y atacar al segundo (o intervenirlo a hurtadillas). Si me cogen desprevenido, tal vez tengan éxito, pero si *entiendo* que la preponderancia de mi deseo de no ser vanidoso es causada por el deseo de torta, quizá pueda resistirme a ellos. En otra oportunidad, mi deseo de una gratificación de corto plazo y mi deseo de largo plazo de espontaneidad pueden constituir una alianza contra mi deseo de mediano plazo de autocontrol. Cuando más de dos motivos intervienen en la elección entre dos alternativas, la idea de «fuerza de la motivación» puede ser indeterminada mientras no sepamos qué alianza se formará.

La Bruyère, el moralista francés del siglo XVII, resumió del siguiente modo dos formas de conflicto motivacional: «Nada es más fácil para la pasión que vencer a la razón; su mayor triunfo es conquistar el interés». Hemos visto que cuando la pasión «vence a la razón», puede, con todo, querer tener a ésta de su lado. Aunque san Pablo dijo: «Puesto que no hago el bien que quiero, sino el mal que no quiero», una reacción más corriente tal vez consista en convencerse de la bondad o la justicia de aquello que, en las garras de la pasión, uno quiere hacer. Cuando la pasión «conquista el interés», puede hacerlo de una de dos maneras. El agente, debido a la *urgencia* que es característica de la emoción (Capítulo 8), quizá no se tome el tiempo de averiguar dónde está su interés. Alternativamente, la fuerza de la emoción tal vez sea tan grande que impulse al agente a actuar *a sabiendas* contra su interés. Ese comportamiento puede equivaler a la debilidad de la voluntad (Capítulo 6).

* * *

NOTA BIBLIOGRÁFICA

George Loewenstein, «Out of control: visceral influences on behavior», *Organizational Behavior and Human Decision Processes*, 65(3), 1996, págs. 272-292, propone una teoría de las motivaciones viscerales. El cálculo del «excedente de accidentes automovilísticos» luego del 11 de septiembre de 2001 está en Gerd Gigerenzer, «Dread risk, September 11, and fatal traffic accidents», *Psychological Science*, 15(4), 2004, págs. 286-287. La

falta de un exceso similar de accidentes en España se documenta en Alejandro López-Rousseau, «Avoiding the death risk of avoiding a dread risk: the aftermath of March 11 in Spain», *Psychological Science*, 16(6), 2005, págs. 426-428. La tricotomía interés-razón-pasión es analizada en Albert Hirschman, *The Passions and the Interests: Political Arguments for Capitalism before Its Triumph*, Princeton, Princeton University Press, 1977 [*Las pasiones y los intereses: argumentos políticos en favor del capitalismo previo a su triunfo*, Barcelona, Península, 1999]; Morton White, *Philosophy*, The Federalist*, and the Constitution*, Nueva York, Oxford University Press, 1987, y en mi *Alchemies of the Mind: Rationality and the Emotions*, Cambridge y Nueva York, Cambridge University Press, 1999 [*Alquimias de la mente: la racionalidad y las emociones*, Barcelona, Paidós, 2002]. George Ainslie, *Picoeconomics: The Strategic Interaction of Sucessive Motivational States within the Person*, Cambridge y Nueva York, Cambridge University Press, 1992, propone los mecanismos faltantes en las intuiciones de Freud. Un estudio clásico del contraste entre la expulsión y la atracción es Diego Gambetta, *Were They Pushed or Did They Jump?: An Analysis of Educational Decisions in North-West Italy*, Cambridge, Cambridge University Press, 1982. Tomo los argumentos de Dión Crisóstomo de Paul Veyne, *L'Empire grécoromain*, París, Seuil, 2005, pág. 217. El principio de la ética judía que menciono se examina en David Daube, *Collaboration with Tyranny in Rabbinic Law*, Londres y Nueva York, Oxford University Press, 1965, y, del mismo autor, *Appeasement or Resistance and Other Essays on New Testament Judaism*, Berkeley, University of California Press, 1987. Desarrollo la idea de estados que son esencialmente subproductos en el Capítulo 2 de *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge y Nueva York, Cambridge University Press, 1983 [*Uvas amargas: sobre la subversión de la racionalidad*, Barcelona, Península, 1988], y la aplico a la cuestión de la redención en «Redemption for wrongdoing: the fate of collaborators after 1945», *Journal of Conflict Resolution*, 50(3), 2006, págs. 324-338, y a la cuestión de la salvación en «Motivations and beliefs in suicide missions», en Diego Gambetta (comp.), *Making Sense of Suicide Missions*, Oxford y Nueva York, Oxford University Press, 2005, págs. 233-258. Véase también Lee Ross y Richard E. Nisbett, *The Person and the Situation: Perspectives of Social Psychology*, Filadelfia, Temple University Press, 1991, págs. 230-232. Analizo la «deferencia a la razón» en los juicios a colaboracionistas

luego de la Segunda Guerra Mundial en el Capítulo 8 de *Closing the Books: Transitional Justice in Historical Perspective*, Cambridge y Nueva York, Cambridge University Press, 2004 [*Rendición de cuentas: la justicia transicional en perspectiva histórica*, Buenos Aires, Katz, 2006]. Se encontrarán pruebas del cambio de los valores atribuidos a diversas características de las decisiones alternativas en Aaron J. Brownstein, «Biased predecision processing», *Psychological Bulletin*, 129(4), 2003, págs. 545-568, y Jack Brehm, «Postdecision changes in the desirability of alternatives», *Journal of Abnormal and Social Psychology*, 52(3), 1956, págs. 384-389.

Notas al pie

[5] Esa actitud también podría deberse, sin embargo, a que su presupuesto no les permite mantener su nivel de consumo (Capítulo 9).

[6] En un caso que ha sido objeto de extensos estudios (véase el Capítulo 20), el rechazo entraña también el sufrimiento de quien ha hecho la oferta, un factor que realzará el motivo del resentimiento.

[7] Para combinar dos de las metáforas de Freud, el yo es como un jinete sobre un caballo indócil (el ello) que es al mismo tiempo cabalgado por un íncubo (el superyó).

[8] Es un hecho que también hay un *riesgo* acumulativo. La posibilidad de que las relaciones sexuales sin protección tengan consecuencias no deseadas tal vez sea pequeña en cualquier oportunidad determinada, pero a lo largo de la vida el riesgo podría ser considerable. En un viaje cualquiera, la posibilidad de sufrir heridas en un accidente automovilístico a raíz de no tener puesto el cinturón de seguridad es pequeña, pero la probabilidad en el transcurso de toda una vida es de alrededor de uno en tres.

[9] El paréntesis que encierra el «parezcan justificar» refleja el posible funcionamiento de la «psicología de la tiranía» (Capítulo 2). Un dilema clásico de la disuasión es que el *odio* que inspira puede terminar por contrarrestar con mucho el *miedo* que pretende causar.

[10] Con respecto a las reglas de la comida *kosher*, de las cuales la prohibición de comer un cabrito hervido en la leche de su madre es sólo una ilustración histórica, durante un tiempo se creyó que tenían su fundamento en razones de higiene. Hasta donde he podido averiguar, hoy se las recomienda con el argumento de que para uno es bueno hacer algo difícil e inútil a la vez. Esta idea parece encarnar la falacia de los subproductos que examino más adelante: el comportamiento que se justifica *meramente* por sus efectos de forja del carácter no tendrá siquiera dichos efectos. No sé, sin embargo, cuántos de quienes siguen la regla lo hacen por esa razón.

[11] Algunas personas envidiosas, por supuesto, no tienen esos escrúpulos. Tal vez vivan en una sociedad donde la envidia no sea vergonzosa, o carezcan sencillamente de vergüenza.

[12] Podemos señalar, para referirnos a ello más adelante (Capítulo 23), que el pensamiento contradesiderativo no puede tener ningún valor de consumo. Por lo tanto, cuando ocurre, es probable que actúe como premisa del comportamiento.

[13] En el volumen final de *En busca del tiempo perdido*, Proust, en una probable reflexión sobre su propia vida, señala que la vana búsqueda de la felicidad puede, no obstante, conducir a intuiciones sobre la condición humana que tal vez «traigan una especie de alegría». La búsqueda de estados que son esencialmente subproductos puede generarlos en forma indirecta, como ocurre cuando un niño que ordena a alguien que se ría hace soltar a éste una carcajada ante tan descabellada exigencia.

[14] La razón por la que se lo liberó fue probablemente que la Resistencia necesitaba el dinero y más adelante se vio obligada a mantener la promesa tácita de inmunidad implicada por la aceptación del cheque.

[15] Esta formulación presupone que el coste de mudarse actúa en un pie de igualdad con los beneficios de haberse mudado, como determinantes de la utilidad global de la mudanza. No obstante, los costes de ésta también pueden actuar como *coacciones* sobre las decisiones. Si la tarifa transatlántica más barata cuesta más que el monto máximo que un italiano pobre puede ahorrar y pedir prestado, este último se quedará en Italia, por mucho mejor que pudiera irle en los Estados Unidos (Capítulo 9).

[16] Esquilo, por ejemplo, escribió sus obras para su representación en una competencia teatral. Cuando el joven Sófocles lo derrotó, sintió tanto pesar que dejó Atenas para marcharse a Sicilia.

[17] Podría, al menos en principio, usarse un compromiso probabilista, mediante el establecimiento de una lotería con valores asignados a las opciones, utilizando la fuerza de las razones en su favor

como valores. Algunas instituciones asignan recursos escasos sobre esa base, pero no he encontrado ningún caso de conflicto motivacional individual que se resolviera de este modo.

[18] Como se señaló, al rechazar el ofrecimiento tal vez logre también que el ofertante no obtenga nada.

[19] En el Capítulo 6 examino los fenómenos más desconcertantes del «perdedor toma todo» que se observan en la debilidad de la voluntad.

[20] La idea de metamotivación no tiene relación con el concepto de metapreferencia. Un ejemplo de este último sería una persona que tuviera dos ordenamientos *diferentes* de preferencias, uno en que la comida está por encima de la dieta y otro en que la dieta está por encima de la comida, y una metapreferencia favorable al segundo. Siguiendo la idea de La Bruyère de que «los hombres son muy vanos, pero nada odian tanto como pasar por serlo», una metamotivación podría equivaler a preferir la preferencia de las dietas sobre la comida, por motivos de salud, a tener *la misma* preferencia por motivos de vanidad.

[21] Según veremos en el Capítulo 12, los agentes también pueden mostrar una deferencia a veces excesiva a la *racionalidad*.

[22] También se produce el fenómeno inverso, en el cual la mente consciente trata de encontrar una razón para una decisión a la que ha llegado la mente inconsciente. Un ejemplo simple es el comportamiento inducido por la hipnosis. Considérense asimismo las explicaciones fabricadas con que dan los amantes para hacer frecuentes llamadas telefónicas al objeto de su afecto. Esta forma de autointerpretación o automalinterpretación no debe confundirse con la trasmutación, que se produce antes y no después de la acción.

[23] En el Capítulo 11 argumento que una persona que tuviera un breve horizonte temporal se negaría, por razones un tanto similares, a tomar una «píldora del descuento» que lo llevara a atribuir mayor importancia a las consecuencias futuras de las acciones presentes. El principio general ilustrado por estas dos píldoras es que una persona racional no querría hacer en dos pasos lo que no querría hacer en uno. Podría querer, claro está, hacer en dos pasos lo que no pudiera hacer en uno.

Capítulo 5
Egoísmo y altruismo

Motivación y comportamiento

El contraste entre las motivaciones egoístas y altruistas es engañosamente simple. Como primera aproximación, entendamos una *motivación altruista* como el deseo de mejorar el bienestar de otros aun al precio de una pérdida neta de bienestar en uno mismo, y un *acto altruista* como una acción que tiene en una motivación altruista una razón suficiente. Si veo que das dinero a un mendigo en la calle, califico ese acto de altruista porque es una acción que *podría* surgir de motivaciones altruistas, sea así o no en los hechos concretos.

Para ver un ejemplo más complejo, consideremos los descubrimientos experimentales concernientes al «castigo altruista», examinados antes en varias oportunidades. En esos estudios, un sujeto a tiene la opción de castigar a otro sujeto B por su comportamiento no cooperativo, con algún coste para sí mismo. No hay interacción cara a cara y los dos sujetos nunca volverán a encontrarse. Pese a ello, muchos sujetos escogen la opción punitiva y de ese modo hacen que B sea más cooperativo en sus posteriores tratos con un tercero, C. El castigo *podría* surgir de motivaciones altruistas, si A prevé y es motivado por el beneficio que su castigo de B reporta a C. En realidad, es más probable que esté motivado por un deseo de venganza.

Hay muchos ejemplos de ese comportamiento fuera del laboratorio. En la Francia del siglo XVIII, los campesinos solían dar respuesta positiva a los pedidos de comida y alojamiento que les hacían mendigos y vagabundos. Si

un campesino se negaba, corría el peligro de que le talaran los árboles, le mutilaran los animales y le incendiaran la casa, actos de destrucción que no redundaban en beneficio alguno para los mendigos y los ponían en riesgo de ser capturados. Aunque no hay razón para creer que los movía de hecho un deseo de hacer que el campesino aceptara en el futuro a otros mendigos, esa motivación sería suficiente para explicarlos. En la Inglaterra preindustrial, las rebeliones campesinas fracasaban de manera invariable en lo referido a sus objetivos inmediatos, y sus dirigentes solían recibir un duro castigo. No obstante, en virtud de su valor de perjuicio, esas rebeliones tenían éxito a largo plazo, pues lograban que las clases propietarias se comportaran más moderadamente de lo que lo habrían hecho en otras circunstancias.

La razón para definir las motivaciones altruistas desde el punto de vista del sacrificio del bienestar y no de los bienes materiales, es excluir casos como el siguiente. Si pago cien mil dólares por la educación universitaria de mi hijo, puede que lo haga porque el bienestar de éste está tan atado al mío que el «sacrificio» hace que ambos estemos mejor.[24] La motivación, aunque referida a terceros, no es altruista.[25] Habría un caso de auténtico altruismo si yo enviara a mi hijo a una escuela pública a pesar de estar en perfectas condiciones de solventar una escuela privada y de creer que ésta sería mejor para él. Al obrar de ese modo, sacrificaría no sólo el bienestar de mi hijo sino también el mío. De manera similar, será más probable que la donación a un banco de sangre (diferenciada de la donación de sangre a un pariente cercano) surja de motivos auténticamente altruistas. En la práctica, sin embargo, tal vez sea imposible decir si una motivación es altruista o simplemente está referida a terceros.

El hecho de que a algunas personas *les guste dar*, porque esa actitud las hace sentirse bien («efecto de la satisfacción moral» [«*warm glow effect*»]), plantea una complicación adicional. Si la satisfacción moral representa la *razón* por la cual ayudan o dan a otros, no sería muy apropiado calificarlas de altruistas. Dan porque, a fin de cuentas, el dar las hace estar mejor. Esto no significa decir que un altruista no pueda sentir también satisfacción moral; sencillamente, ésta no se incluiría entre sus motivos (inconscientes) para ayudar o dar. Insistamos, sin embargo, en que en la práctica quizá sea imposible trazar esta distinción. Me extenderé sobre ello en el Capítulo 15.

Sean cuales fueren los problemas de identificar las motivaciones altruistas, hay pruebas abundantes del *comportamiento* altruista. La

Carnegie Foundation entrega habitualmente medallas a individuos que han salvado la vida de otros con gran riesgo para sí mismos. Muchas personas dan sangre sin que les paguen por el esfuerzo.[26] En Noruega, la mayoría de los riñones para trasplante son donados por parientes del receptor. La extracción del riñón implica un riesgo médico, pero no hay ninguna recompensa monetaria.[27] Muchos individuos, sobre todo mujeres, cuidan a sus padres ancianos además de trabajar y ocuparse de sus propias familias. En numerosos países, más de la mitad de la población adulta hace donaciones regulares de dinero con propósitos de beneficencia. Luego del maremoto de 2004, se observaron elevados picos de donaciones en muchos países desarrollados. En tiempos de guerra, algunos individuos tratan de disimular sus discapacidades a fin de que se les permita combatir. Numerosos soldados se ofrecen como voluntarios para cumplir misiones peligrosas (y algunos lo hacen incluso para participar en misiones suicidas). Cuando la gente vota en elecciones nacionales y contribuye así a la viabilidad de la democracia, incurre en algunos costes y no obtiene virtualmente ningún beneficio privado. La lista podría extenderse indefinidamente.

La razón por la cual no podemos inferir del comportamiento altruista la existencia de motivaciones altruistas es que otras motivaciones pueden *remedar el altruismo*. Conforme a la terminología del Capítulo 4, podemos ver el altruismo como una especie de *razón*, susceptible de ser efectivamente simulada sea por *interés* o por *pasión*. (Las palabras «remedar» o «simular» tal vez den a entender, aunque no deben hacerlo necesariamente, un esfuerzo consciente por engañar a otros acerca de nuestra verdadera motivación.) Mucha gente que se preocupa poco por ser desinteresada se preocupa mucho por recibir elogios a causa de su desinterés. Así, Hume se equivocaba, sin duda, al afirmar que «amar la gloria de las obras virtuosas es una prueba *segura* del amor por la virtud» (las bastardillas son mías). Montaigne, en contraste, aseveraba: «Cuanto más llamativa es una buena acción, más rebajo de su bondad la sospecha que en mí nace de que haya sido realizada más por ser llamativa que por ser buena; al ser exhibida, casi se vende». En el límite, los únicos actos virtuosos son los que nunca salen a la luz. La angelical abuela del narrador de Proust había internalizado este principio de manera tan exhaustiva que atribuía todas sus buenas acciones a motivos egoístas. Habida cuenta de que

la virtud tiene esa característica de borrarse a sí misma, puede haber más de la que salta a la vista. Por otras razones, desde luego, puede haber menos.

Aprobatividad y vergüenza

Montaigne también reconocía la escasez de la virtud cuando establecía una distinción entre «monedas» motivacionales buenas y falsas: actuar en función de lo que está bien y actuar en función de lo que otras personas creen de uno. Como la primera motivación es poco habitual, los legisladores tal vez tengan que apoyarse en la segunda:

> Si esa falsa idea [la preocupación por lo que otras personas piensan] sirve a las instituciones para mantener a los hombres en su deber […], acreciéntese con osadía y aliméntese entre nosotros cuanto sea posible. […] Puesto que los hombres, por su incapacidad, no pueden pagarse con moneda buena, empléese para ello la falsa. Fue practicado este medio por todos los legisladores y no hay sociedad en la que no exista cierta mezcla de vanidad ceremoniosa o de opinión mentirosa que sirva de brida para mantener al pueblo en el deber.

Napoleón hizo eco a esta idea cuando, al defender la creación de la Légion d'Honneur en 1802, dijo que «tales fruslerías mueven a los hombres». (Sus viejos soldados del ejército republicano reaccionaron vigorosamente contra ese invento.) La *aprobatividad* (el deseo de que los otros piensen bien de uno) es una moneda falsa que puede tener que sustituir la moneda buena del altruismo y la moral. De manera alternativa, la *vergüenza* (el deseo de que los otros no piensen mal de uno) puede actuar como moneda falsa. Las normas sociales tal vez induzcan a las personas a abstenerse de acciones que, de lo contrario, podrían realizar. El respeto por la norma no basta, sin embargo, para que los otros piensen bien de ellas. La aprobación queda reservada para los actos *supererogatorios*, es decir, los que van más allá de la norma. Lo que es obligatorio en una sociedad puede ser supererogatorio en otra. En Noruega y los Estados Unidos hay una norma social (moderada) por la cual un hermano debe donar un riñón si se lo necesita (y es apto) para un trasplante,[28] mientras que en Francia ese comportamiento podría verse como supererogatorio. En ciertos círculos sociales, las donaciones a la beneficencia son imperativas.

Estas motivaciones pueden ilustrarse con dos ejemplos contrastantes de la política del siglo XVIII. En la primera Asamblea Constituyente francesa (1789-1791), los diputados sacrificaron varias veces intereses importantes, en una actitud que los llevó desde renunciar a sus privilegios feudales hasta

declararse inelegibles para la primera legislatura ordinaria. Si bien sus motivaciones eran complejas, un componente importante era el deseo de que los vieran como desinteresados. En palabras del biógrafo de uno de ellos, estaban «embriagados de desinterés». Más o menos en la misma época, en los Estados Unidos, George Washington manifestó reiteradas veces su temor de que otros pensaran que lo motivaba el interés privado. (Al mismo tiempo, era consciente de que una preocupación excesiva por la propia virtud podía parecer poco virtuosa.) Para mayor ilustración, examinemos dos concepciones del honor. De acuerdo con una, éste debe *adquirirse* por medio de hazañas gloriosas. De acuerdo con otra, se lo supone como un umbral, pero puede *perderse* a causa de hechos vergonzosos.

Que la aprobatividad o la vergüenza puedan remedar el altruismo dependerá de los criterios sustantivos que otros apliquen en la evaluación del comportamiento. Algunas sociedades pueden atribuir alto valor a virtudes que no muestran tendencia sistemática alguna a remedar el altruismo, y cuya expresión pueden estimular con esa atribución. El deseo de honor tal vez induzca toda clase de comportamientos socialmente dispendiosos. Las fruslerías de Napoleón pretendían alentar a los soldados a arriesgar la vida para mayor gloria de Francia, no promover el bienestar de los franceses. Algunos individuos pueden escoger una vida abnegada debido al elogio que su sociedad hace de los virtuosos religiosos, pero cenobitas y monjes suelen concentrarse más en los rituales del culto que en sus semejantes. Que yo sepa, las comunidades que asignan mucho valor a la educación y el aprendizaje no tienden a generar comportamientos más altruistas que otras. En las sociedades occidentales modernas, el culto de la belleza estimula un comportamiento egocéntrico que parecería ser enemigo del interés por los otros. En las sociedades sujetas a lo que se ha dado en llamar «familismo amoral» (el sur de Italia se ha mencionado como ejemplo), hay normas sociales que desaconsejan ayudar a los extraños en apuros o cumplir la ley. Es difícil, por lo tanto, decir en líneas generales si el deseo de recibir elogios o de evitar la censura tiende a remedar el altruismo.

Trasmutaciones

Como se discutió en el capítulo anterior, la motivación original del agente también puede *trasmutarse* de interés en razón. El mecanismo subyacente a esta alquimia es el *amour propre* o *amor propio*, el deseo de estima y autoestima. En tanto que la aprobatividad y la vergüenza, que derivan del deseo de estima, sólo afectan el comportamiento exterior, el deseo de autoestima puede afectar las motivaciones internas. La mayoría de la gente no quiere verse como si sólo la moviera su interés personal. Aun cuando actúe para satisfacer ese interés, trata de dar una apariencia desinteresada a sus acciones. En las décadas anteriores a la Guerra de Secesión de los Estados Unidos de América, la esclavitud dejó de ser en el viejo sur una cuestión de mero interés y se convirtió en una *causa*, defendida con argumentos principistas. Éste es un tipo muy corriente de ideología política. Como señaló Marx, «no vaya nadie a formarse la idea limitada de que la pequeña burguesía quiere imponer, por principio, un interés egoísta de clase. Ella cree, por el contrario, que las condiciones *particulares* de su emancipación son las condiciones *generales* fuera de las cuales no puede ser salvada la sociedad moderna y evitada la lucha de clases».

Aunque pueda ser arduo decir si la apariencia desinteresada es sincera o hipócrita, sería un error creer que habitualmente es lo último. En los esfuerzos por autojustificar su comportamiento, la gente tiene, de hecho, dos grados de libertad. Por un lado, hay muchas *teorías causales de fachada verosímil* que pueden utilizarse como respaldo de la idea de que las acciones que benefician a uno mismo también beneficiarán a otros. Por ejemplo, las teorías del crecimiento económico «por goteo» dan a entender que los impuestos bajos aplicados a los ricos también beneficiarán a los pobres. Por otro lado, hay tantas *concepciones normativas* de la justicia, la equidad o el bien común *que parecen plausibles*, que una persona tendría que ser desafortunada o incompetente para no encontrar una que (de conformidad con alguna teoría causal de fachada verosímil) no coincidiera con su interés propio. En el Capítulo 4 señalé que la gente podría escoger entre una filosofía de la beneficencia basada en la equidad y otra basada en la utilidad para justificar las bajas donaciones.[29] En tales casos, la gente se inclina en forma espontánea e inconsciente por una combinación de teoría causal y concepción normativa que pueda justificar un comportamiento conforme a su interés propio. Aunque no entendemos cómo sucede, sabemos que sucede.

No hace falta que el «remedo» del altruismo se apoye en la hipocresía o la trasmutación. El razonamiento egoísta puede generar un comportamiento altruista por medio de mecanismos absolutamente admisibles como la elección *detrás del velo de ignorancia* o la reciprocidad. En el Capítulo 4 mencioné, como ejemplo de la primera, el argumento de George Mason de que el interés a largo plazo de las familias debía inducir una preocupación por el bienestar de todas las clases de la sociedad.[30] Argumentos similares pueden ser válidos en la vida de un solo individuo. En las sociedades que tienen un bajo índice de desempleo pero que atraviesan un período de rápido cambio estructural, la mayoría quizá vote a un partido que propone elevados beneficios para los desempleados si cree que su trabajo está entre los que podrían desaparecer. Lo que parece solidaridad podría ser únicamente una forma de seguro.

Reciprocidad

La reciprocidad puede ser una simple relación diádica, como cuando cada una de las partes de una relación actual enfrenta la alternativa de cooperar o no cooperar. Un granjero tal vez coseche en agosto y otro en septiembre, y cada uno de ellos puede beneficiarse con la ayuda del otro. Si el granjero cuya cosecha se hace antes solicita la ayuda del otro pero luego niega la suya en septiembre, es improbable que reciba asistencia en agosto del año siguiente. De haber cooperación, es probable que se desarrolle una relación estable de asistencia mutua que, aunque no se apoye en sentimientos de camaradería, puede alentarlos. Durante la Primera Guerra Mundial, algunas tropas alemanas y británicas establecieron una tregua tácita, una actitud de vivir y dejar vivir consistente en cañonear al enemigo con menos agresividad de la que podrían haber dado muestra.[31] También en este caso apareció con el tiempo una actitud amistosa hacia el otro bando, pero como *resultado* de la cooperación y no como su causa.

Descartes describió una forma más compleja de reciprocidad multilateral o indirecta:

> La razón que me impulsa a creer que quienes no hacen nada, salvo por su propia utilidad, también deberían, si desean ser prudentes, trabajar, como hacen otros, por el bien de los demás, y procurar complacer a todos tanto como les fuera posible, es que uno ve de ordinario que aquellos a quienes se considera serviciales y prontos a agradar también reciben una cantidad de buenas acciones de otros, aun de personas que nunca han tenido obligaciones con ellos; y no recibirían

> estas cosas si la gente los creyera de otro temperamento; y los afanes que ponen en complacer a otras personas no son tan grandes como el provecho que la amistad de quienes los conocen les proporciona. Puesto que otros sólo esperan de nuestra parte los actos que podemos realizar sin inconveniencia para nosotros mismos, tampoco nosotros esperamos más de ellos; a menudo sucede, no obstante, que actos que cuestan poco a otros nos son de mucho provecho, y pueden incluso salvarnos la vida. Es cierto que ocasionalmente uno dilapida su afán en hacer el bien y que, por otra parte, de vez en cuando gana al hacer el mal; pero eso no puede modificar la regla de prudencia que se relaciona únicamente con las cosas que ocurren más a menudo.

En la reciprocidad directa, A ayuda a B si y sólo si B ha ayudado a A. En la reciprocidad indirecta, A ayuda a B si B ha ayudado a C. Como veremos en capítulos siguientes, una distinción similar es válida para la «reciprocidad negativa»: A tal vez dañe a B si B ha dañado a A, pero también si B ha dañado a C. La existencia de la reciprocidad indirecta sugiere que las personas podrían actuar de manera altruista a fin de granjearse una *reputación* por sus motivaciones altruistas. Otras personas deberán entonces decidir si el comportamiento refleja un auténtico altruismo o no es más que un deseo estratégico de construir una reputación de altruista (Capítulo 20). En este caso, la reputación se valora en el terreno instrumental, no por sus características intrínsecas. Mientras que la aprobatividad lleva al agente a desear ser estimado por la estima misma, la reputación se busca por las recompensas materiales que podría significar.

La gente también puede actuar con reciprocidad en situaciones únicas que no brindan la oportunidad de una recompensa ulterior. Si a se comporta de manera altruista con B, B puede responder con la misma moneda aun cuando ambos sepan que no habrá más interacciones entre ellos. El granjero que cosecha en agosto podría ayudar al que cosecha en septiembre aunque planeara emigrar antes de la siguiente estación. Es posible, desde luego, imaginar razones egoístas de esa reciprocidad. Quizás el granjero que levanta la cosecha antes teme que el otro lo castigue de alguna manera si no muestra reciprocidad, o que terceros de cuya asistencia depende lo aíslen. En condiciones experimentales, sin embargo, esos efectos pueden excluirse. En los juegos experimentales que examinaremos más adelante (Capítulo 15 y Capítulo 20), los sujetos interactúan en forma anónima a través de terminales de ordenadores, lo cual excluye todo efecto cara a cara como la vergüenza o la turbación. Con frecuencia, los juegos también están diseñados para que una persona interactúe sólo una vez con otro de los participantes.

Aun en esas rigurosas condiciones se observa que hay reciprocidad. En un juego de confianza, un jugador, el «inversor», tiene la opción de

transferir a otro jugador, el «operador», cualquier suma de su dotación de diez unidades monetarias. El experimentador triplica luego la cifra entregada, de modo que si el inversor transfiere diez, el operador recibe treinta. Este último puede decidir a su turno devolver al primero cualquier monto, desde cero hasta la suma total aumentada (tres veces lo que le ha enviado el inversor). En un experimento, los inversores transfieren en promedio alrededor de dos terceras partes de su asignación a los operadores, mientras que éstos hacen en promedio un reembolso un poco más grande. Cuanto mayor sea la transferencia «inicial», mayor será el reembolso. Estos descubrimientos son compatibles con una serie de supuestos motivacionales, *excepto* la hipótesis de que ambos agentes están motivados por el egoísmo material y saben que el otro tiene esa misma motivación. Según dicha hipótesis, el inversor, que espera un reembolso igual a cero, hará una transferencia también igual a cero. Como este resultado no se observa, deben de actuar motivaciones referidas a otros o «preferencias sociales». Éstas pueden no estar a la altura del altruismo o la equidad, dado que en algunos experimentos los operadores no devuelven más de lo que han recibido, pese a lo cual este monto es más grande que el que impondría el egoísmo.

Normas morales, sociales y cuasimorales

En varios de los próximos capítulos volveré a las implicaciones de estos experimentos y de otros relacionados con ellos. Ahora me limitaré a trazar distinciones entre tres tipos de motivaciones «referidas a otros». Las *normas morales* incluyen la de ayudar a otros en apuros, la del reparto equitativo y la del «kantismo cotidiano» (haz lo que sería óptimo si todo el mundo hiciera lo mismo). Las *normas sociales* (Capítulo 22) incluyen las normas formales, de venganza y de regulación del uso del dinero. Las normas que califico de *cuasimorales* comprenden la norma de la reciprocidad (ayuda a quienes te ayudan y daña a quienes te dañan) y la norma de la cooperación condicional (coopera si los otros lo hacen, pero de lo contrario no lo hagas). Tanto las normas sociales como las normas cuasimorales son condicionales, en cuanto se activan con la presencia o el comportamiento de otras personas. Mi hipótesis es que las normas sociales se activan cuando otras personas pueden observar lo que el agente está haciendo, y las normas cuasimorales, cuando el agente puede observar lo

que hacen otros.[32] Las normas morales, en contraste, son incondicionales. Lo que nos indican hacer puede depender, a buen seguro, de lo que otros hagan. Si baso mi filosofía de la beneficencia en la utilidad, la cantidad de bien que haga (y, por ende, la cantidad que dé) dependerá de cuánto den los otros. La norma misma, con todo, no hace referencia a otros donantes; sólo se refiere a los receptores.

Dos casos de respuestas individuales a la escasez de agua ilustrarán la distinción entre las normas sociales y las normas cuasimorales. En Bogotá, durante la imaginativa gestión del alcalde Antanas Mockus, la gente siguió una norma cuasimoral al reducir su consumo de agua. Aunque el control individual no era factible, el consumo agregado de agua de la ciudad se mostraba en televisión, de modo que cada uno de sus habitantes podía saber si, en general, los demás cumplían. Al parecer, la cantidad de personas que lo hacían era suficiente para sostener la cooperación condicional. La gente se decía: «Como los demás están reduciendo su consumo, es justo que yo también lo haga». En contraste, cuando hay escasez de agua en California, son las normas sociales las que parecen llevar a los residentes a limitar su consumo. El consumo externo, por ejemplo para regar el césped, puede ser controlado no sólo por los vecinos sino también por inspectores municipales. El control del consumo interno puede quedar a cargo de los visitantes, que tal vez (y sin «tal vez») expresen su desaprobación si el inodoro está limpio.[33] De hecho, también en Bogotá había un control del comportamiento individual, puesto que los niños a veces hacían pasar un mal rato a sus padres si éstos no economizaban agua.[34]

Las normas cuasimorales pueden ser, como es obvio, de suma eficacia para inducir comportamientos altruistas. ¿Se limitan a *remedar* el altruismo, o *son* motivaciones altruistas? La razón por la cual las califico de cuasimorales y no morales explica también por qué me inclino por la primera respuesta. La norma de la reciprocidad nos autoriza a *no* ayudar a otros que están en apuros a menos que éstos nos hayan ayudado anteriormente. Una norma moral típica es la que indica que debe ayudárselos en forma incondicional, aun cuando no haya un historial previo de asistencia. La norma de la cooperación condicional nos permite utilizar cantidades normales de agua si ninguna otra persona reduce el consumo, mientras que tanto el utilitarismo como el kantismo cotidiano adherirían a la reducción unilateral. Las normas morales, podríamos decir, son *proactivas*; las normas cuasimorales sólo son *reactivas*. Otra manera de expresar la

diferencia consiste en decir que el sentimiento de injusticia parece tener mayor fuerza motivacional que el sentido de justicia. Como veremos más adelante (Capítulo 20), las propuestas que los participantes en un experimento tienden a rechazar como injustas, con la consecuencia de que ni ellos ni los proponentes consiguen nada, son del mismo orden de magnitud que las que los proponentes tienden a ofrecer cuando no los limita el temor al rechazo.

Al parecer, podríamos identificar el funcionamiento de motivos auténticamente altruistas si se cumplieran dos condiciones. En primer lugar, la acción que beneficia a otros es proactiva, no reactiva. En segundo lugar, es anónima, en el sentido de que la identidad del actor benéfico no es conocida ni por el beneficiario ni por terceros.[35] Podemos imaginar, por ejemplo, a una persona que envía una donación anónima a una entidad benéfica como Oxfam o pone dinero en la alcancía de una iglesia vacía. El segundo ejemplo no es tan definido como uno quisiera, dado que la persona podría estar motivada por su creencia en que Dios la observa y la recompensará. La creencia tal vez sea ilógica (un ejemplo de la «falacia de los subproductos») pero, aun así, podría ser bastante común. El primer ejemplo quizá parezca más inequívoco. No obstante, aun los actos más puros de altruismo, como las donaciones anónimas a extraños, pueden surgir de motivos oscuros. Según Kant,

> En realidad, es absolutamente imposible determinar por medio de la experiencia y con absoluta certeza un solo caso en que la máxima de una acción, por lo demás conforme con el deber, haya tenido su asiento en fundamentos exclusivamente morales y por medio de la representación del deber. Pues a veces se da el caso de que, a pesar del examen más penetrante, no encontramos nada que haya podido ser bastante poderoso, independientemente del fundamento moral del deber, como para mover a tal o cual buena acción o a un gran sacrificio, sólo que de ello no podemos concluir con seguridad que la verdadera causa determinante de la voluntad no haya sido en realidad algún impulso secreto del egoísmo oculto tras el simple espejismo de aquella idea: solemos preciarnos mucho de poseer algún fundamento determinante lleno de nobleza, pero es algo que nos atribuimos falsamente. Sea como fuere, y aun ejercitando el más riguroso de los exámenes, no podemos nunca llegar por completo a los más recónditos motores de la acción, puesto que cuando se trata del valor moral no importan las acciones, que se ven, sino sus principios íntimos, que no se ven.

Kant dice que aun cuando no estemos representando ante un público exterior, nunca podemos saber si estamos actuando ante la *audiencia interna*. El acto de ocultar la propia virtud, que Montaigne consideraba tan virtuoso, no puede ocultarse a uno mismo. Como señaló La Rochefoucauld, el *amour propre* «siempre encuentra compensaciones, y aun cuando

renuncia a la vanidad no pierde nada». Y como también dijo: «Si el amor puro existe, libre de la escoria de nuestras otras pasiones, está oculto en las profundidades de nuestro corazón y es desconocido aun para nosotros mismos». En el mejor de los casos, dijo Proust, tal vez podamos conocer nuestros verdaderos motivos gracias a los otros: «no nos es dado conocer más que las pasiones ajenas, y lo que llegamos a conocer de las nuestras lo sabemos por los demás. Nuestras pasiones no accionan sobre nosotros más que en segundo lugar, por medio de la imaginación, que coloca en lugar de los móviles primeros, móviles de relevo que son más decentes».

La atribución de motivaciones

Además de la motivación propia del agente, la explicación de su comportamiento debe recurrir a menudo a sus creencias sobre las motivaciones de los otros. Al formar esas creencias, el agente se enfrenta al mismo dilema hermenéutico que el historiador o el especialista en ciencias sociales. Puesto que le resulta imposible tomar tal como vienen las motivaciones profesadas de los otros, puede utilizar triangulaciones del tipo general que examiné en el Capítulo 3. Puede valerse, además, de técnicas que sólo se aplican a las interacciones cara a cara. Otras personas quizá sean capaces de identificar a un mentiroso aficionado por su lenguaje corporal (o la falta de éste), dado que la concentración en lo que dice lo lleva a pasar por alto los gestos que por lo común acompañan el discurso espontáneo. Por otra parte, para verificar motivos profesados, uno puede poner una trampa al agente. Mientras que los historiadores no están en condiciones de hacerlo con los individuos que estudian, y los especialistas en ciencias sociales suelen tener razones éticas que lo impiden, un empleador, un cónyuge o un padre tal vez se sientan menos limitados.

Muchas veces, la atribución de motivos a los otros está teñida de malicia. Dada la alternativa entre creer que una acción altruista tiene por causa una motivación altruista y creer que está basada en el egoísmo, a menudo suponemos lo último, aun cuando no haya fundamentos positivos para hacerlo. Si bien esa desconfianza puede tener sentido por razones prudenciales (Capítulo 26), en muchos casos esta justificación no está a nuestro alcance. El chisme, por ejemplo, parece estar con frecuencia motivado por lo que los moralistas franceses, siguiendo a Agustín, llamaron *malignidad* y *debilidad* de la naturaleza humana.[36] De acuerdo con La

Rochefoucauld, «si no tuviéramos faltas no nos regocijaríamos tanto en las faltas de los otros». De hecho, como también escribió este mismo autor, nuestro deseo de encontrar faltas en los otros es tan fuerte que a menudo nos ayuda a encontrarlas: «Nuestros enemigos están más cerca de la verdad en su opinión de nosotros que nosotros mismos». No obstante, aun cuando nuestros enemigos estén más cerca de la verdad, también yerran, aunque yerren menos, en la dirección contraria. En una escala de cero a diez, si yo soy seis, creeré que soy nueve, y mis enemigos me asignarán un cuatro.

Para un análisis de esta actitud, a veces denominada «hermenéutica de la sospecha», no puedo hacer nada mejor que citar a Jeremy Bentham (traducido de su desmañado francés):

> Cualquiera que sea la postura que el rey [Luis XVI] adopte, cualquiera que sea el sacrificio que haga, nunca logrará silenciar a estos difamadores: ellos son un gusano que el mal temperamento y la vanidad jamás dejarán de alimentar incluso en el más sano de los cuerpos políticos. La más prolífica fuente de esta injusticia es en primerísimo lugar la vanidad. Uno quiere ocuparse sutilmente de todo […] y prefiere la más artificiosa suposición a la vergüenza de haber sospechado que la conducta de una persona pública podría tener un motivo laudable. Si Washington persiste en su retiro, éste sólo puede ser un medio de utilizar el camino a la anarquía para allanar el camino al despotismo. Si Necker, en vez de aceptar como cualquier otro un pago por sus servicios, paga de su propio peculio por tener el permiso de prestarlos, la actitud sólo puede ser una manera refinada de satisfacer su codicia. Si Luis XVI abdica el poder legislativo en favor de su pueblo, sólo puede ser como resultado de un elaborado plan para recuperarlo y, aún más, en un momento favorable.

Una ironía es que la última de las acusaciones especiosas mencionadas en este texto (escrito a principios de 1789) estaba probablemente justificada en el otoño de 1790. Uno de los asesores más íntimos del rey, Saint-Priest, escribió que por esa época éste había dejado de resistir las usurpaciones de la legislatura porque «se había convencido de que la Asamblea se desacreditaría por sus propios errores». Las teorías conspirativas pueden ser exactas, porque las conspiraciones existen. No obstante, la tendencia a encontrarlas quizá deba menos a la experiencia que a una renuencia perniciosa a admitir que las figuras públicas podrían actuar por buenas razones.

* * *

NOTA BIBLIOGRÁFICA

Este capítulo se basa en mi «Altruistic behavior and altruistic motivations», en Serge-Christophe Kolm y Jean Mercier Ythier (comps.),

Handbook of the Economics of Giving, Reciprocity and Altruism, Ámsterdam, Elsevier, 2006, págs. 183-206. Otros capítulos de este volumen, sobre todo el artículo de introducción de Serge-Christophe Kolm, proporcionan abundante información empírica y análisis teóricos. La referencia a los mendigos franceses es de Georges Lefebvre, *La Grande peur de 1789*, París, Armand Colin, 1988, pág. 40 [*El gran pánico de 1789: la Revolución francesa y los campesinos*, Barcelona, Paidós, 1986], mientras que la referencia a las rebeliones campesinas inglesas se ha tomado de Edward P. Thompson, «The moral economy of the English crowd in the 18th century», *Past and Present*, 50(1), febrero de 1971, págs. 76-136 [«La economía moral de la multitud en la Inglaterra del siglo XVIII», en *Tradición, revuelta y conciencia de clase*, Barcelona, Crítica, 1984, págs. 62-134]. Se encontrará un análisis del altruismo de la «satisfacción moral» en James Andreoni, «Impure altruism and donations to public goods: a theory of warm-glow giving», *Economic Journal*, 100(401), 1990, págs. 464-477. Sobre las actitudes con respecto a la donación de riñones, véase Hilde Lorenzen y Florence Paterson, «Donations from the living: are the French and Norwegians altruistic?», en Jon Elster y Nicolas Herpin (comps.), *The Ethics of Medical Choice*, Londres, Pinter, 1994, págs. 10-115 [«Las donaciones en vida: ¿son altruistas los franceses y los noruegos?», en *La ética de las decisiones médicas*, Barcelona, Gedisa, 2000]. Tomo la idea (y la palabra) de «aprobatividad» de Arthur O. Lovejoy, *Reflections on Human Nature*, Baltimore, Johns Hopkins Press, 1961 [*Reflexiones sobre la naturaleza humana*, México, Herrero Hermanos, 1965]. El papel del desinterés en la Revolución francesa se examina en Barry M. Shapiro, «Self-sacrifice, self-interest, or self-defense? The constituent assembly and the "self-denying ordinance" of May 1791», *French Historical Studies*, 25(4), 2002, págs. 625-656. Sobre el paralelo estadounidense, véase Gordon Wood, «Interest and disinterestedness in the making of the constitution», en Richard Beeman, Stephen Botein y Edward C. Carter II (comps.), *Beyond Confederation: Origins of the Constitution and American National Identity*, Chapel Hill, University of North Carolina Press, 1987. En lo relacionado con el mecanismo de la trasmutación, véase mi *Alchemies of the Mind: Rationality and the Emotions*, Cambridge y Nueva York, Cambridge University Press, 1999, Capítulo 5 [*Alquimias de la mente: la racionalidad y las emociones*, Barcelona, Paidós, 2002]. El ejemplo de «uno por otro» de la Primera Guerra Mundial proviene de

Robert M. Axelrod, *The Evolution of Cooperation*, Nueva York, Basic Books, 1984 [*La evolución de la cooperación: el dilema del prisionero y la teoría de los juegos*, Madrid, Alianza, 1986]. Sobre el juego de la confianza, véase Colin Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction*, Nueva York, Russell Sage, 2004, Capítulo 2.7. Sobre la detección de mentiras, véase Paul Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*, Nueva York, Norton, 1992 [*Cómo detectar mentiras: una guía para utilizar en el trabajo, la política y la pareja*, Barcelona, Paidós, 1999]. El pasaje de Jeremy Bentham se extrajo de su *Rights, Representation, and Reform: Nonsense upon Stilts and Other Writings on the French Revolution*, Oxford y Nueva York, Oxford University Press, 2002, págs. 17-18.

Notas al pie

[24] También podría haber pagado los aranceles escolares aun cuando fueran tan altos que el gasto hubiera empeorado mi situación en términos de bienestar. En ese caso, el pago de aranceles inferiores se explica por motivos referidos a terceros que están por encima del altruismo.

[25] Cuando Marx escribió que la sociedad comunista estaría *más allá* del egoísmo y el altruismo, en vez de estar basada en esta última motivación, es posible que haya querido decir que cualquier sacrificio *material* que cada uno hiciera por los demás no entrañaría, por similares razones, ningún sacrificio del *bienestar*.

[26] En rigor de verdad, se ha sostenido que la falta de pago es importante para excluir a personas con enfermedades infecciosas que podrían donar sangre para obtener dinero.

[27] En la mayoría de los países, en realidad, la venta de riñones de personas vivas es ilegal. En este caso, la motivación subyacente a la ley puede consistir tanto en proteger a los individuos indigentes contra sí mismos como proteger a los receptores contra el trasplante de partes corporales de baja calidad.

[28] En los Estados Unidos, los médicos ayudan con frecuencia a un potencial donante a resistir esa presión, diciéndole en una etapa temprana del proceso que, si alguien les solicita la donación, están dispuestos a proporcionarle una excusa médica para no hacerlo.

[29] En este caso, la teoría causal que justifica donar poco si los otros dan mucho reside en que las donaciones tienen una utilidad marginal decreciente para los receptores. La teoría basada en la equidad que justifica las donaciones pequeñas si los otros dan poco no necesita una premisa causal.

[30] El gobernador Morris propuso un argumento similar en el debate sobre la representación de los estados en el Senado: «El apego a los estados y la importancia de éstos han sido el veneno de este país. No podemos eliminarlo; pero acaso podamos arrancar los dientes a la serpiente. Quien habla desea que nuestras ideas se amplíen al verdadero interés del hombre, en vez de quedar circunscritas dentro del estrecho círculo de un lugar determinado. Y después de todo, qué pequeño puede ser el motivo esgrimido por el egoísmo para defender esa actitud. ¿Quién puede decir que él mismo, y mucho menos sus hijos, será el año próximo un habitante de este o aquel estado?» En otra oportunidad, el gobernador Morris se situó en el extremo receptor del mismo argumento. En respuesta a la adhesión de Elbridge Gerry a la propuesta de Morris de limitar la representación de los futuros estados del oeste, Roger Sherman replicó: «Proveemos para nuestra posteridad, para nuestros hijos y nuestros nietos, que serán con igual probabilidad ciudadanos de los nuevos estados del oeste como de los viejos estados. Sobre la base de esta sola consideración, no deberíamos hacer una discriminación como la propuesta por la moción».

[31] Aunque el alto mando trató de detener la práctica, ésta era difícil de controlar.

[32] Ambas pueden reforzarse mutuamente, cuando el agente está en condiciones de observar lo que hacen los observadores mismos. Si veo que estás arrojando papeles al suelo, tal vez no me importe que me veas haciendo otro tanto. Sin embargo, si veo que guardas cuidadosamente la envoltura de tu helado en el bolsillo, la equidad y el temor a la desaprobación pueden llegar a combinarse para generar conformidad (véase también el Capítulo 22).

[33] El ahorro de agua también es una preocupación en épocas normales. Para lograrlo, la ciudad de Nueva York tiene leyes que establecen el volumen máximo de los tanques de los inodoros. En gran parte de Europa, hay inodoros con dos botones que lanzan diferentes cantidades de agua para diferentes usos. Este último sistema es interesante en cuanto no se basa ni en oportunidades ni en incentivos (Capítulo 9), sino únicamente en la buena voluntad inobservable de la persona.

[34] Descubrimientos experimentales también apuntan a este mecanismo. En una campaña de ahorro de energía, se colocaron carteles en duchas de vestuarios, que instaban a los estudiantes a cerrar el grifo mientras se enjabonaban y a volver a abrirlo únicamente para enjuagarse. El efecto de

los carteles fue mínimo. Sin embargo, cuando uno o dos sujetos, participantes cómplices del experimento, comenzaron a cumplir las indicaciones, el acatamiento de los demás usuarios de las duchas se incrementó en forma espectacular. Si bien los sujetos cómplices no dijeron nada a los demás, su comportamiento podría actuar como un reproche tácito a quienes no cumplían las instrucciones.

[35] En los experimentos, el experimentador desconoce la identidad del sujeto. En las donaciones a la beneficencia, la identidad del donante es desconocida para los funcionarios de la institución benéfica.

[36] Discrepo de quienes quieren explicarlo por su papel en la imposición de normas sociales. Es cierto, el chisme puede actuar como un multiplicador sobre las sanciones informales que sostienen las normas sociales, pero creo que su origen tiene raíces más profundas.

Capítulo 6
Miopía y previsión

Más allá del ascenso de gradiente

El principio de placer de Freud (Capítulo 4) es la tendencia a buscar una satisfacción inmediata de los deseos. Una manifestación de esta tendencia es la adopción de aquella creencia que a uno le gustaría que fuera cierta, en vez de la creencia respaldada por pruebas. El pensamiento desiderativo me hace sentir bien aquí y ahora, aun cuando después pueda hacerme dar de narices en el suelo. Otra de sus manifestaciones es la que se produce en la elección entre dos acciones que inducen diferentes corrientes de utilidad temporal. El principio de placer impone la elección de la corriente que tenga la mayor utilidad en el primer período, con prescindencia de la forma que tomen las corrientes en períodos posteriores.

Dicho de un modo más general, quien decide, trátese de una lombriz de tierra o de una empresa, puede dedicarse al *ascenso de gradiente*. En un momento dado, explora las opciones *cercanas* para ver si alguna de ellas rinde beneficios *inmediatos* más grandes que el statu quo. La limitación a las opciones cercanas es una forma de «miopía espacial»: ojos que no ven, corazón que no siente. La limitación a los beneficios inmediatos es una forma de miopía temporal: el principio de placer. La lombriz de tierra explora el entorno para ver si algún punto cercano es más húmedo que el que ocupa en este momento, y se desplaza hacia él si lo encuentra. La empresa explora el «espacio» de rutinas que están próximas a lo que hace actualmente para encontrar una que prometa un mejor desempeño a corto

plazo y, de hallarla, la adopta. Tras un tiempo, la lombriz o la empresa tal vez lleguen a permanecer en un lugar que es superior (a corto plazo) a todas las posiciones cercanas. Han alcanzado un *máximo local*.

Los seres humanos pueden hacer algo mejor. La intencionalidad (la aptitud de re-presentar lo ausente) nos permite ir más allá del principio de placer y tomar en cuenta las consecuencias temporalmente remotas de las elecciones presentes. La planificación de lo venidero nos permite hacer elecciones con mejores consecuencias que las derivadas de decisiones tomadas minuto a minuto o segundo a segundo. En algunos casos, esas acciones previsoras pueden emprenderse para dar mejor satisfacción a necesidades actuales, como ocurre cuando un alcohólico se abstiene de beber en un restaurante de las cercanías para poder comprar al mismo precio una botella entera en una tienda situada en un lugar distante. En otros casos, las acciones se emprenden para satisfacer necesidades futuras, como cuando ahorro para mi vejez. Mientras que el primer tipo de previsión también se observa en animales no humanos, habitualmente se ha interpretado que el segundo tipo de previsión está más allá de la capacidad de éstos. Sin embargo, algunas pruebas recientes sugieren que los primates quizá sean capaces de planificar sobre la base de necesidades esperadas y no actuales. Sea como fuere, obrar sobre la base de necesidades proyectadas es obviamente una operación más sofisticada.

Desearía dar cuatro ejemplos de la acción sobre la base de consecuencias temporalmente remotas. Los primeros tres ejemplos también se analizan en capítulos posteriores.

RECULER POUR MIEUX SAUTER. Esta frase francesa, equivalente aproximado de «un paso atrás, dos pasos adelante», tiene su ilustración en un hecho fundamental de la vida económica, a saber, que, con el fin de invertir para tener un mayor consumo en el futuro, es preciso consumir menos en el presente. El agente acepta un estado que es inferior al statu quo porque es una condición para realizar una alternativa superior más adelante. No hace falta decir que esto sólo tiene sentido si 1) el estado inferior permite al agente sobrevivir, y 2) las ganancias en el estado superior son lo bastante grandes para justificar la pérdida que implica pasar al estado inferior.

ESPERAR. Muchos vinos, aunque ya buenos en el momento de embotellarlos, mejoran con el tiempo. Para beneficiarse con este hecho, el agente tiene que estar dispuesto a rechazar una opción (beber el vino ya mismo) que es superior al statu quo, porque el rechazo es una condición para realizar más adelante un resultado aún mejor. Insistamos en que la postergación del consumo quizá no siempre tenga sentido; por ejemplo, si el agente no espera vivir lo suficiente para disfrutar del vino añejado. Si se busca un ejemplo de mayores consecuencias, considérese la elección de un cónyuge. En vez de proponer matrimonio o aceptar una propuesta matrimonial en la primera oportunidad que aparece un candidato aceptable, uno podría esperar a alguien aún más adecuado. El riesgo, ilustrado con abundancia en la literatura mundial, es que no aparezca nadie más adecuado.

DISPARAR POR DELANTE DEL BLANCO. Para acertar en un blanco móvil, es menester apuntar no al lugar donde está, sino donde estará en el momento del encuentro. De manera similar, para perseguir un blanco móvil, se debe apuntar en línea recta al lugar donde estará, en vez de seguir la trayectoria curva inducida por el hecho de apuntar siempre a su posición actual.

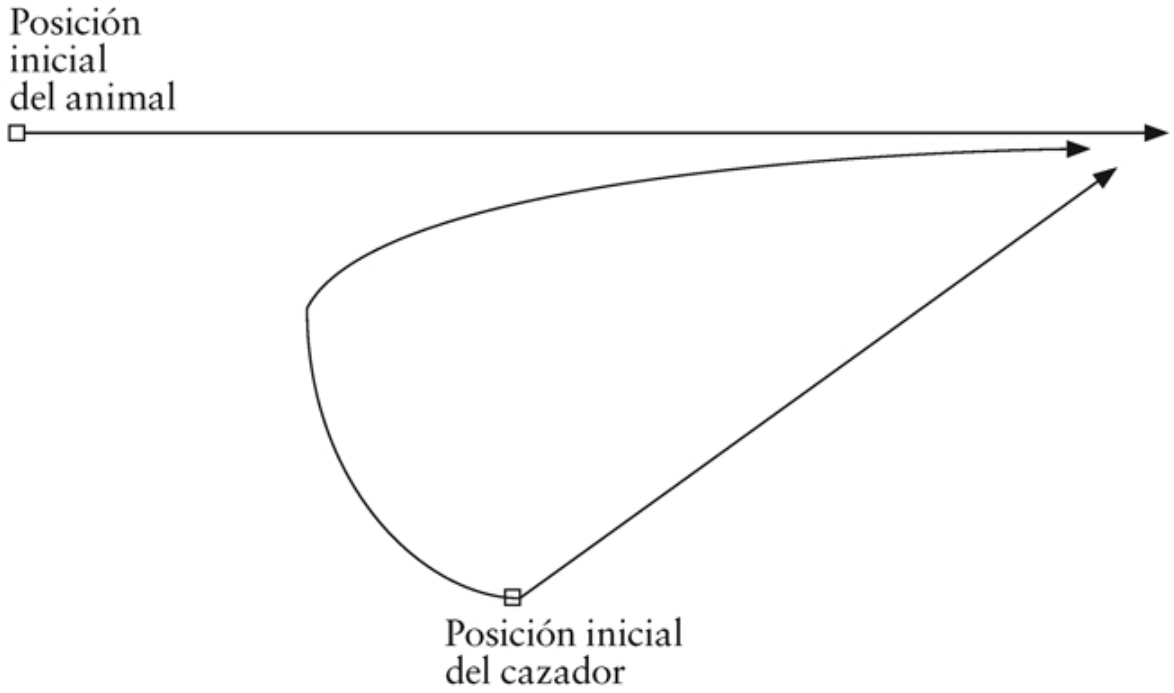


FIGURA 6.1

En la Figura 6.1, el cazador, aun cuando se mueva un poco más lentamente que el animal, puede darle alcance si va en línea recta hacia el punto donde estará en algún momento calculable del futuro. Sin embargo, si siempre apunta en dirección a la posición actual del animal, siguiendo la trayectoria curva del diagrama, nunca le dará alcance. Como veremos (Capítulo 17), la selección natural en un medio ambiente cambiante puede verse de acuerdo con esta perspectiva.

UNA LÍNEA RECTA NO SIEMPRE ES EL CAMINO MÁS CORTO ENTRE DOS PUNTOS. Cuando tratamos de alcanzar un blanco estacionario, una línea recta no siempre es la trayectoria más eficiente. En la Figura 6.2, el rescatista podría correr impulsivamente en línea recta hacia el nadador en aprietos hasta llegar a la orilla, y luego nadar la distancia restante. Sin embargo, si se hubiera tomado tiempo (¡pero no demasiado!) para reflexionar, tal vez habría advertido que, como puede ir más rápido corriendo que a nado, alcanzaría antes al nadador si tomara una trayectoria indirecta que, aunque más larga en su conjunto, tiene un trecho más corto en el agua. Nos comportamos de este modo cuando tomamos una autopista en vez de la carretera que, en el mapa, parece ser más corta. En la planificación económica, ese «comportamiento de autopista» es con frecuencia el óptimo.

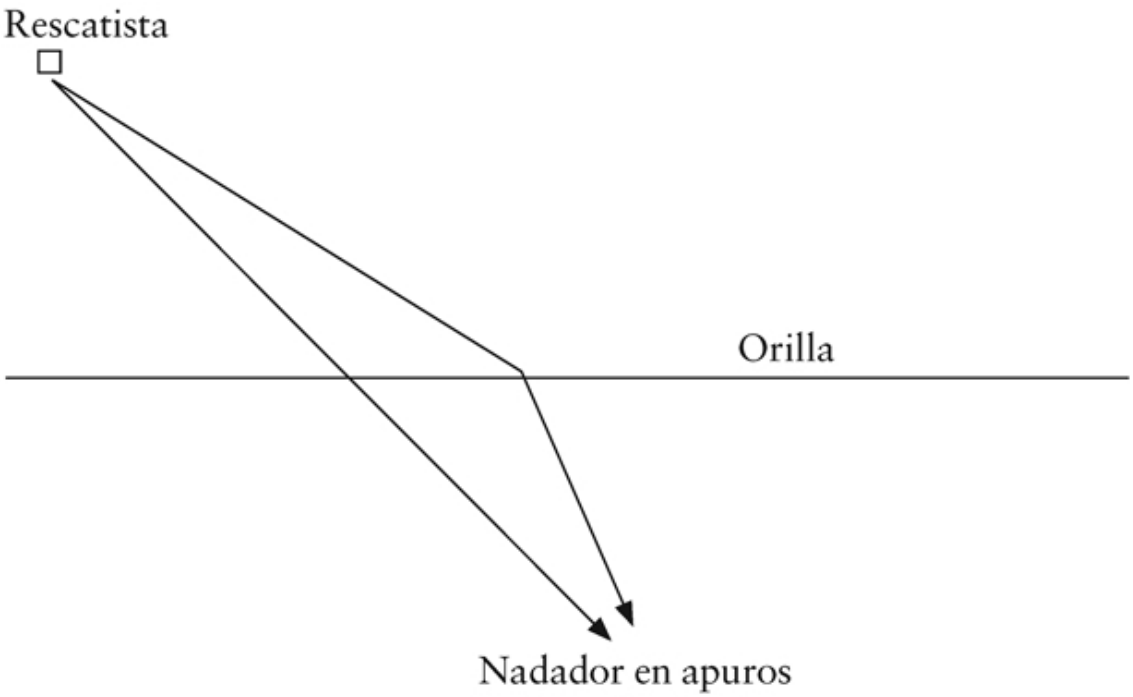


FIGURA 6.2

Descuento temporal

La existencia de la capacidad de planificación a largo plazo no implica que se la haya de utilizar. Para que las consecuencias percibidas de largo plazo signifiquen algo para el comportamiento presente, los agentes deben estar *motivados* a tomarlas en cuenta. En el lenguaje de los psicólogos, deben estar dispuestos a *postergar la satisfacción*. En el lenguaje de los economistas, no deben estar sometidos a un *descuento temporal* excesivo.[37] Son necesarios los elementos cognitivos y motivacionales. Si los resultados futuros están envueltos en la incertidumbre, no pueden motivar el comportamiento presente. Si implican riesgo, su fuerza motivadora también se atenúa. La aptitud de los resultados futuros para dar forma al comportamiento presente se ve afectada tanto por *el tiempo en el cual* y *la probabilidad según la cual* ocurrirán. Los mecanismos a través de los que afectan la elección son, respectivamente, el descuento temporal y las actitudes de riesgo.

Como sugiere la expresión, el descuento (o miopía) temporal es la tendencia a atribuir menos importancia a las recompensas en el futuro distante que a las recompensas en el futuro cercano o el presente.[38] Si se les plantea una alternativa entre tener cien dólares hoy y ciento diez dólares dentro de un año, la mayoría de las personas preferirá la primera. Esta preferencia podría tener, sin embargo, una serie de orígenes.

> Algunos podrían preferir la recompensa inmediata porque pueden invertir los fondos y retirar más de ciento diez dólares al cabo de un año. Otros podrían aceptar los cien dólares hoy porque necesitan el dinero para sobrevivir. La obtención de una suma mayor más adelante no tiene valor si esperan estar muertos por entonces. Supongamos asimismo que debo elegir entre pescar peces con las manos en la corriente y hacer una red que me permita pescar muchos más. Sin embargo, como no puedo pescar mientras hago la red, el coste de oportunidad de fabricar ésta quizá sea tan alto que no pueda darme ese lujo.
> Otros más podrían tomar la recompensa menor porque tienen una enfermedad que entraña un riesgo del 10% de morir dentro de un año. En términos más generales, cuando planificamos para el futuro, debemos tomar en cuenta el hecho de que sabemos que moriremos, pero no cuándo. Si la suma futura es una recompensa esperada, que implica una posibilidad del 50% de que sean ciento treinta dólares y una posibilidad del 50% de que sean noventa dólares, la aversión del riesgo podría inducir una preferencia por aceptar hoy los cien dólares seguros.
> Para terminar, algunas personas podrían preferir la recompensa anticipada simplemente porque llega antes. Esta actitud es una muestra de *descuento temporal puro*. Así como una casa grande vista a la distancia parece más pequeña que una casa pequeña en primer plano, una gran suma en el futuro puede parecer, subjetivamente, más pequeña que una suma pequeña en el presente. En lo que sigue, consideraré sólo este caso.

¿El descuento temporal puro es irracional? Supongamos que una persona hace un nutrido descuento de las recompensas futuras. En vez de recibir educación universitaria, que conlleva un sacrificio temporario de ingresos para obtener más adelante un ingreso más alto, acepta un trabajo de bajo nivel con escasas probabilidades de ascenso inmediatamente después de terminar la escuela secundaria. Como ignora el impacto de largo plazo del hábito de fumar y de los alimentos con colesterol elevado, tiene una corta esperanza de vida.[39] Si no respeta la ley por razones morales, las consideraciones prudenciales no la disuadirán de violarla. En otras palabras, es muy probable que su vida sea corta y desdichada. Si esto no es comportamiento irracional, ¿qué es?

A mi juicio, el descuento temporal puro no es irracional de por sí. Tal vez haga que la vida del agente sea peor que si éste se preocupara por el futuro, pero esto también puede suceder con las motivaciones egoístas. Quien sólo se preocupa por sí mismo quizá termine por tener una vida triste y empobrecida, pero esa razón no debe llevarnos a decir que el egoísmo es

irracional. Me ocupo de estas cuestiones en el Capítulo 11. Aquí me concentro en la manera adecuada de conceptualizar el descuento temporal. Tenemos a nuestra disposición varios enfoques, con implicaciones radicalmente diferentes.

Para modelizar el descuento temporal, los teóricos de la decisión supusieron tradicionalmente que la gente descuenta la utilidad futura *de manera exponencial*. Una unidad de períodos t de utilidad en el futuro tiene un valor presente de k^t, donde $k < 1$ es el factor de descuento por período. El descuento exponencial tiene, desde un punto de vista normativo, el atractivo de permitir una *planificación coherente*. Si una corriente de recompensas tiene un valor presente más grande que otra en un momento, lo tendrá en todos los demás. Por ello, el agente nunca está sujeto a una inversión de las preferencias, que (en ausencia de razones para cambiar de opinión) se considera de ordinario una marca distintiva de la irracionalidad.

En un plano empírico, sin embargo, la noción de planificación coherente tiene menos sentido. La observación causal muestra, y la observación sistemática confirma, que casi todos estamos frecuentemente sujetos a una inversión de las preferencias. A menudo omitimos llevar a la práctica las intenciones de ahorrar, hacer ejercicios a la mañana, practicar piano, cumplir con nuestras citas, etc. Tal vez llame a mi dentista el 1º de marzo y concierte una cita para el 1º de abril, para no hacer otra cosa que volver a llamar y cancelarla el 30 de marzo, con el pretexto (falso) de que debo ir a un funeral. Para explicar estas variedades de irracionalidad cotidiana (y muchos otros fenómenos), podemos reemplazar el supuesto del descuento exponencial por el del *descuento hiperbólico*.

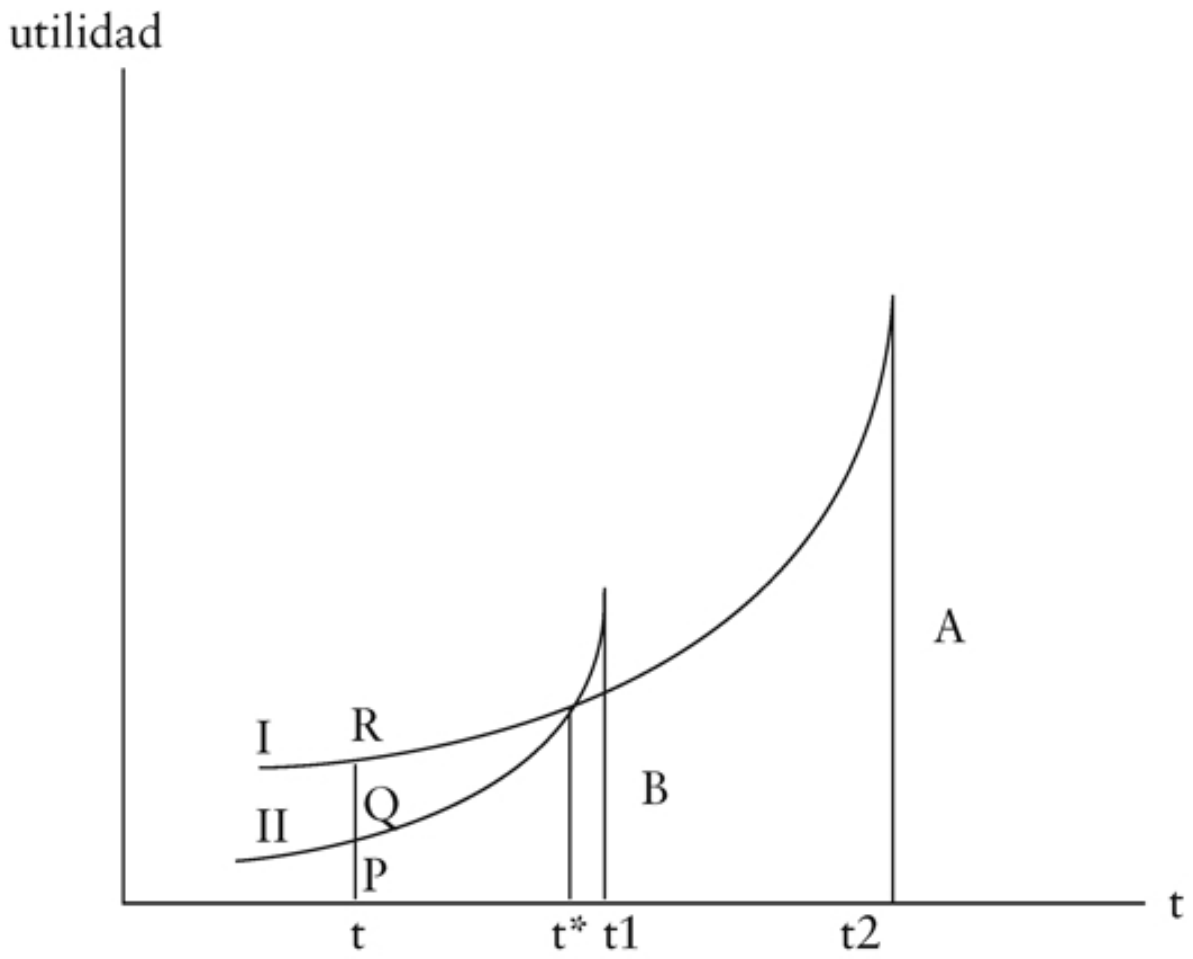


FIGURA 6.3

Supongamos que el valor descontado presente de una unidad de períodos *t* de utilidad en el futuro es igual a $1/(1 + kt)$. (En el ejemplo de abajo supongo que $k = 1$, pero en el caso más general, *k* podría ser cualquier número positivo: cuanto más grande sea, menos se preocupará el agente por el futuro.) Supongamos, por otra parte, que en $t = 0$ el agente enfrenta la elección entre una recompensa de 10 en $t = 5$ y una recompensa de 30 en $t = 10$. En $t = 0$, el valor presente de la primera es 1,67, y el de la segunda es 2,73. Un agente que maximice el valor presente se hará la intención de elegir la recompensa diferida. En $t = 1$, el valor presente de la primera recompensa es 2, y el de la segunda es 3. En $t = 2$, los valores son 2,5 y 3,3, respectivamente; en $t = 3$, son 3,3 y 3,75, y en $t = 4$, ascienden a 5 y 4,29. Es decir que en algún momento entre $t = 3$ y $t = 4$ la primera recompensa deja de ser la opción menos elegida y se convierte en la más elegida, *como*

resultado del mero paso del tiempo, y nada más. Es fácil ver, de hecho, que el giro se produce en $t = 3{,}5$, que es cuando llamo a mi dentista para cancelar la cita.

Este patrón es aún más fácil de ver en un diagrama. En la Figura 6.3, el agente puede elegir o bien la recompensa pequeña B en $t1$ o esperar hasta $t2$ y obtener la recompensa A, más grande. Las curvas hiperbólicas I y II representan los valores presentes de esas recompensas conforme se los evalúa en diversos momentos anteriores. Se trata, de hecho, de *curvas de indiferencia* (Capítulo 9) que representan la compensación entre el momento en que una recompensa está disponible y las dimensiones de ésta. En el tiempo t, por ejemplo, al agente le resulta indiferente recibir de inmediato la recompensa PQ o recibir la recompensa pequeña en $t1$, y también le es indiferente hacerse enseguida con PR u obtener la recompensa grande en $t2$. Como en el tiempo t el valor presente de A es más grande que el de B, el agente se formará la intención de escoger A. No obstante, debido a que las curvas hiperbólicas se cruzan en t^*, en ese momento se produce una inversión de las preferencias y el agente elige B.[40]

La apuesta de Pascal

Podemos utilizar la apuesta de Pascal para ilustrar la relación entre los descuentos temporales exponencial e hiperbólico. Pascal quería persuadir a aquellos de sus amigos que eran jugadores librepensadores de que apostaran a la existencia de Dios, dado que aun la más mínima posibilidad de bienaventuranza eterna compensaría los placeres terrenales más grandes. Su argumento esconde muchas complejidades, de algunas de las cuales nos ocuparemos en el próximo capítulo. Aquí sólo quiero aludir a una cuestión que el filósofo no menciona: ¿el valor presente (descontado) de la bienaventuranza eterna es finito o infinito? Si es finito, el jugador podría preferir gozar de sus placeres en la Tierra en vez de esperar hasta la vida después de la muerte.

Supongamos, para simplificar, que cada período en la vida después de la muerte proporciona una unidad de utilidad experimentada; que la persona espera morir en n años a partir de hoy, y, por último, que descuenta exponencialmente el bienestar futuro por un factor de $k (0 < k < 1)$. Si Dios existe y le concede la salvación sobre la base de su fe, el valor presente de la bienaventuranza en el primer año luego de su muerte es k^n unidades de

utilidad, el del segundo año es k^{n+}, y así sucesivamente. Por una cuestión de álgebra elemental, esta suma infinita ($k^n + k^{n+} + k^{n+} \ldots$) resulta en una suma finita $k^n/(1 - k)$. Es concebible al menos que esta suma sea inferior al valor presente de n años de vida hedonista en la Tierra. En contraste, si el agente está sujeto al descuento hiperbólico, la suma infinita $1/(n + 1) + 1/(n + 2) + 1/(n + 3) \ldots$ se incrementa más allá de cualquier valor finito dado, lo cual implica que, si comparamos los valores presentes, en última instancia la bienaventuranza de la salvación dará alcance a cualquier valor terrenal. Aun cuando esa bienaventuranza se multiplique por una pequeña probabilidad (tan pequeña como uno quiera) de que Dios exista, el producto aumentará más allá de cualquier número finito.

Supongamos, sin embargo, que el interlocutor de Pascal tiene oportunidades habituales de jugar. Cuando considera las cosas por anticipado, prefiere asistir a misa antes que jugar, porque la primera actividad, a la larga, lo llevará a creer y le asegurará una expectativa de bienaventuranza infinita. Por la lógica del descuento hiperbólico, sin embargo, la inminencia de la oportunidad de jugar inducirá una inversión de preferencias. La persona de marras se formará la intención de jugar una sola vez más para luego empezar a ir a misa. Dirá, con san Agustín, «dame castidad y continencia, pero no todavía». A la semana siguiente, será válido el mismo razonamiento. Así, la estructura misma del descuento temporal que garantiza que la bienaventuranza eterna tenga el mayor valor presente, también impedirá al jugador dar los pasos necesarios para alcanzarla.

Debilidad de la voluntad

Como muestra este ejemplo, el descuento hiperbólico puede echar luz sobre el problema clásico de la debilidad de la voluntad (DV). Una persona de voluntad débil se caracteriza del siguiente modo:

1. La persona tiene una razón para hacer X.
2. La persona tiene una razón para hacer Y.
3. En opinión de la persona, la razón para hacer X tiene mayor peso que la razón para hacer Y.
4. La persona hace Y.

Con frecuencia se sostiene que las emociones, en particular, tienen la capacidad de inducir una acción contra la opinión misma del agente. Cuando Medea, en la obra de Eurípides, está por matar a sus hijos, dice: «Sé en verdad qué mal pretendo hacer. Pero más fuerte que todos mis juicios es mi furia». En la versión de la obra escrita por Ovidio, ella dice: «Aunque renuente, una compulsión desconocida me abruma. Apremiada a tomar este camino o aquel […], veo el mejor y lo apruebo, pero sigo el peor».

Estas expresiones, como los cuatro enunciados utilizados para caracterizar la debilidad de voluntad, son ambiguas o poco precisas, en el sentido de que no se menciona *cuándo* deben ser ciertas. Definamos de la siguiente manera una *concepción estricta de la debilidad de la voluntad*:

1. La persona tiene una razón para hacer X.
2. La persona tiene una razón para hacer Y.
3. La persona hace Y, a pesar de juzgar *en el momento de la acción* que la razón para hacer X tiene mayor peso que la razón para hacer Y.

Imaginemos a una persona que ha resuelto dejar de fumar y va a una fiesta donde le ofrecen un cigarrillo. Acepta el ofrecimiento, a sabiendas de que no debería hacerlo. Una persona que está haciendo una dieta tal vez acepte el ofrecimiento de un postre, aunque sepa que no es una buena idea. Si bien esta concepción de la debilidad de la voluntad no tiene nada de imposible, tropieza con dos problemas empíricos. Sería difícil establecer que la acción y la «opinión más pertinente» coexisten en el mismo momento, y no que la segunda ha cambiado una fracción de segundo antes de producirse la primera. Además, que yo sepa, nadie ha especificado el mecanismo causal mediante el cual el deseo de hacer Y adquiere mayor eficacia causal que el deseo de hacer X.

Para eludir estos problemas, podemos definir una *concepción general de la debilidad de la voluntad*, que permita que la opinión del agente de que debe hacer X y su decisión de hacer Y se produzcan en diferentes momentos:

1. La persona tiene una razón para hacer X.
2. La persona tiene una razón para hacer Y.

3. Tras reflexionar con calma, la persona opina que la razón para hacer X tiene mayor peso que la razón para hacer Y.
4. La persona hace Y.

Sócrates negaba que, en sentido estricto, la DV fuese posible. Aristóteles también estuvo cerca de señalar lo mismo. Admitía la DV en el sentido general y mencionaba como ejemplo el caso de una persona cuyo juicio en el momento de la acción estuviese bajo la influencia del alcohol. Supongamos que voy a la fiesta de la oficina, bebo demasiados tragos, ofendo a mi jefe y cortejo a su mujer. En el momento, estas acciones parecen perfectamente naturales. Sin embargo, si alguien me hubiera sugerido con anterioridad que podía actuar de ese modo, yo habría rechazado la idea por ser incompatible con mi juicio calmo y reflexivo. Si hubiese estado convencido de que mi juicio podía ahogarse en el alcohol, habría permanecido lejos. Una vez consumado el hecho, tal vez lamente amargamente mi conducta.

Este caso, mostrado en la Figura 6.4, es un caso de *inversión temporaria de las preferencias*, no de DV en sentido estricto. Hay al menos tres mecanismos que pueden provocar tales cambios. Uno es la *proximidad temporal*, como se ha explicado en el examen del descuento hiperbólico. Otro es la *proximidad espacial*, según la ilustra el fenómeno de la dependencia de una señal. Este mecanismo explica, por ejemplo, muchos casos de recaída entre adictos. Aun después de años de abstinencia, una señal del entorno tradicionalmente asociada al consumo de drogas puede desencadenar la recaída. Tal vez baste con ver la parafernalia de la droga en un programa de televisión. La resolución de iniciar una dieta quizá quede debilitada a la vista de la mesa rodante de los postres. En resumidas cuentas, también en estos casos el agente elige lo que prefiere de acuerdo con su concepción *en el momento de la decisión*. Para terminar, las *pasiones* son capaces de inducir un cambio temporario de preferencias, en virtud del hecho de que suelen tener una vida media breve (Capítulo 8). También pueden inducir una inversión de preferencias al hacer que el agente preste menos atención al futuro remoto.[41]

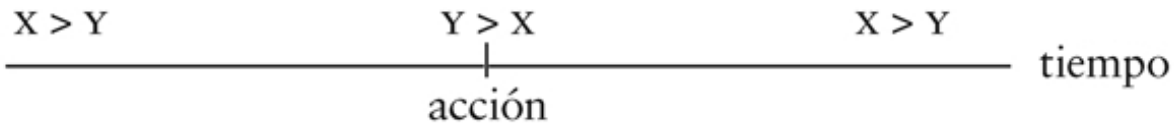


FIGURA 6.4

Podemos ampliar esta idea para incluir los cambios temporarios (y motivados) en las *creencias* del agente. De conformidad con esta concepción muy general, la debilidad de la voluntad puede ser también el resultado del autoengaño (o el pensamiento desiderativo). Tras haber decidido con anterioridad a una fiesta tomar sólo dos tragos a fin de poder conducir el automóvil de regreso a casa sin inconvenientes, una persona podría, bajo la influencia de su deseo de beber un tercer trago, decirse, contra el peso de la evidencia, que eso no modificará sus aptitudes de conductor.[42] Su preferencia (por la conducción segura) se mantiene sin cambios, pero su creencia referida a las condiciones en que puede conducir con seguridad se ha modificado. La persona también podría, desde luego, sufrir un cambio temporario de preferencias, si decide que el hecho de pasarla bien en la fiesta es tan importante que compensa los riesgos (que acaso ella perciba con exactitud) de conducir ebria.

* * *

NOTA BIBLIOGRÁFICA

Se encontrarán pruebas de que los primates pueden ser capaces de planificar en función de necesidades futuras (no experimentadas en el presente) en Nicholas J. Mulcahy y Joseph Call, «Apes save tools for future use», *Science*, 312, 2006, págs. 1038-1040. Dos libros de referencia sobre el descuento temporal y otros aspectos de la elección intertemporal son los de George Loewenstein y Jon Elster (comps.), *Choice over Time*, Nueva York, Russell Sage Foundation, 1992, y George Loewenstein, Daniel Read y Roy Baumeister (comps.), *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, Nueva York, Russell Sage Foundation, 2003. Examino la apuesta de Pascal con mayor detalle en «Pascal and decision theory», en Nicholas Hammond (comp.), *The Cambridge Companion to Pascal*, Cambridge y Nueva York, Cambridge University Press, 2004, págs. 53-72. Las pruebas neurofisiológicas del descuento temporal cuasihiperbólico se hallarán en Samuel McClure y otros, «Separate neural systems evaluate immediate and delayed monetary rewards», *Science*, 306, 2004, págs. 503-507. Los análisis modernos de la debilidad de la voluntad derivan de Donald Davidson, «How is weakness of

the will possible?», en *Essays on Actions and Events*, Nueva York, Oxford University Press, 1980, págs. 21-42 [«¿Cómo es posible la debilidad de la voluntad?», en *Ensayos sobre acciones y sucesos*, Barcelona y México, Crítica/Instituto de Investigaciones Filosóficas de la UNAM, 1995]. Comento sus ideas en «Davidson on weakness of will and self-deception», en Lewis Hahn (comp.), *The Philosophy of Donald Davidson*, Chicago, Open Court, 1999, págs. 425-442. La formación de creencias motivadas se examina en David Francis Pears, *Motivated Irrationality*, Nueva York, Oxford University Press, 1984. Me ocupo con mayor detalle del vínculo entre la debilidad de la voluntad y la inversión de preferencias en «Weakness of will and preference reversal», en Jon Elster y otros (comps.), *Understanding Choice, Explaining Behavior: Essays in Honour of Ole-Jørgen Skog*, Oslo, Oslo Academic Press, 2006.

Notas al pie

[37] En este libro, la frase «una tasa elevada de descuento temporal» significará que las recompensas futuras tienen un pequeño valor presente. La frase «un elevado factor de descuento» significará que tienen un gran valor presente. Para ilustrar y motivar esta terminología aparentemente extraña, supóngase que el agente es indiferente entre tres unidades de recompensa mañana y dos unidades hoy. La recompensa futura se descuenta (reduce) en un tercio. El factor de descuento (el número por el cual tenemos que multiplicar la recompensa futura para obtener su valor presente) es dos tercios.

[38] Algunos individuos, como los avaros patológicos, tal vez adjudiquen más importancia a la utilidad futura que a la utilidad presente. Para ellos, el tiempo de consumir nunca está del todo maduro.

[39] Cincuenta años atrás, muchas personas podrían haber «ignorado» estas consecuencias, en cuanto que no las conocían. Si bien esta situación es menos probable en nuestros días, la gente aún puede «ignorarlas», en el sentido de atribuirles menos importancia en sus decisiones. No pocas veces, es posible que se encuentre también en un estado de «ignorancia motivada» (una forma de pensamiento desiderativo) acerca de las consecuencias.

[40] Hay una manera alternativa y ligeramente diferente de representar el descuento hiperbólico, que se apoya en la idea intuitiva de que la gente traza una distinción radical entre el presente y todos los demás momentos, al atribuir mayor importancia al bienestar en el período actual que en todos los períodos ulteriores. Además, diferencia *entre* esos períodos ulteriores. En un ejemplo de tres períodos, donde u_i es el bienestar experimentado en el período i, el valor presente o la suma descontada de la utilidad es $u + b(du + du)$. Intervienen aquí dos factores de descuento. En comparación con el presente, toda utilidad futura, con prescindencia de cuándo se experimente, es descontada por un factor b. Por otra parte, todas las utilidades futuras se descuentan exponencialmente por un factor d. El momento presente tiene una preponderancia visceral que lo hace destacarse por encima de todos los otros, mientras que los períodos ulteriores pierden poco a poco su poder motivador por algo más afín a una ilusión óptica. Este patrón, llamado «descuento cuasihiperbólico», tiene en común con el descuento hiperbólico propiamente dicho la capacidad de inducir inversiones de preferencias. Y difiere de él en cuanto el valor presente de una corriente infinita de recompensas iguales (como en la apuesta de Pascal) tiene una suma finita. La neurofisiología suministra algunas pruebas de que el descuento cuasihiperbólico, aunque introducido sólo como una aproximación útil al descuento hiperbólico, es de hecho la representación más precisa.

[41] De hecho, la inversión de preferencias causada por las preferencias temporales hiperbólicas puede ser remedada por cambios emocionalmente inducidos en el factor de descuento asociado a las preferencias temporales exponenciales. Supongamos que el agente enfrenta la elección entre dos alternativas, A y B, que ofrecen las respectivas recompensas (2, 5, 6) y (5, 4, 1) en tres períodos sucesivos. Con una tasa de descuento uniperiódica de 0,8 (y una tasa biperiódica de 0,64), los valores presentes de las dos alternativas (tal y como se los evalúa en el primer período) son respectivamente de 9,84 y 8,84. Con una tasa de descuento uniperiódica de 0,6 (y una tasa biperiódica de 0,36), los valores son 7,16 y 7,96. No es de sorprender que el agente deje de preferir la alternativa con las mejores consecuencias de largo plazo cuando las emociones lo llevan a prestar menor atención al futuro.

[42] En contraste, si le preocupa más la posibilidad de que la policía la detenga que la de sufrir un accidente, le resultará más difícil convencerse de que el tercer trago no hará que el contenido de alcohol en la sangre supere el límite legalmente permitido. Como argumento en el capítulo siguiente, aun el pensamiento desiderativo está (un tanto) sometido a las restricciones de la realidad.

Capítulo 7
Creencias

¿Qué es «creer» en algo?

Para entender el papel de las creencias en la generación de la acción, tenemos que entender su naturaleza, sus causas y sus consecuencias. Como señalé en las observaciones de introducción a la segunda parte, no siempre resulta claro qué significa «creer» que algo es cierto, por ejemplo, que hay vida después de la muerte. Muchas de las grandes figuras religiosas se refirieron a su lucha constante con la duda. En sus momentos de credulidad, ¿era su fe tan simple e incondicional como la de alguien que jamás ha conocido la duda? ¿Los seguidores del comunismo que «creían» que el partido no podía obrar mal realmente lo *creían*?[43] ¿Cómo podemos explicitar la diferencia entre el pesimista congénito que tiende a creer lo peor y el decisor prudente que se limita a actuar *como si* el peor escenario fuera cierto?

Por otra parte, en el lenguaje de todos los días «creencia» sugiere algo menos que una adhesión plena. *Creo* que mañana va a llover, pero también sé que podría equivocarme. No me limito a creer que estoy casado; lo *sé*. En los análisis filosóficos, el conocimiento suele definirse como una creencia verdadera justificada, una creencia que mantiene una relación particular tanto con el mundo (es verdadera) como con la masa de pruebas en poder del agente (está justificada). Sin embargo, ninguna de estas dos características del conocimiento capta la certeza subjetiva que a menudo subyace a la frase «ya sé» en el discurso corriente. Esa certeza no es

simplemente el límite de una probabilidad del 97%, el 98%, el 99%, el 99,9%, etc. Es cualitativamente diferente de todo lo que no sea certidumbre.[44]

Este «efecto de certidumbre» aparece en el siguiente experimento. Se pidió a un grupo de sujetos que expresara sus preferencias sobre diversas opciones. (Los números entre paréntesis indican la proporción de sujetos que prefirieron una opción determinada.)

> Un 50% de probabilidades de ganar un recorrido de tres semanas por Inglaterra, Francia e Italia (22%).
> Una recorrida de una semana por Inglaterra, con certeza (78%).
>
> A otro grupo se le plantearon las siguientes alternativas:
>
> Una probabilidad del 5% de ganar una recorrida de tres semanas por Inglaterra, Francia e Italia (67%).
> Una probabilidad del 10% de ganar una recorrida de una semana por Inglaterra (33%).

Los integrantes del primer grupo tienden a preferir la opción «sólo Inglaterra» porque existe la *seguridad* de conseguirla. Una vez que se la reduce a la misma probabilidad que la otra opción, esta última parece más atractiva. Las vacilaciones de los soldados a quienes se pregunta si se ofrecerán como voluntarios en misiones de sumo peligro pueden ser desproporcionadamente menores que las de aquellos cuya participación se solicita para misiones suicidas. Es posible, por supuesto, que los primeros estén también bajo el influjo del pensamiento desiderativo («a mí no me va a pasar»), que no tiene influencia sobre los segundos.

Cuatro actitudes cognitivas

Aun si hacemos a un lado estos problemas, la idea de creencia sigue siendo ambigua. Podemos distinguir entre cuatro actitudes cognitivas hacia el mundo, con fuerza decreciente. En primer lugar está el modo de la *certeza*. En segundo lugar, el modo del *riesgo*, en el cual los agentes asignan probabilidades, basadas en frecuencias pasadas o en su propio juicio, a cada uno de una serie de resultados mutuamente excluyentes y exhaustivos en su conjunto. En tercer lugar tenemos el modo de la *incertidumbre*, en el cual la gente conoce la serie de resultados mutuamente excluyentes y exhaustivos en conjunto, pero se ve en la incapacidad de atribuirles ninguna probabilidad (cardinal).[45] Para terminar, está el modo de

la *ignorancia*, en el que tanto la gama de resultados posibles como su probabilidad de aparición son desconocidos, en parte o por completo. Según las memorables palabras de Donald Rumsfeld, ex secretario de defensa de los Estados Unidos, enfrentamos no sólo cantidades conocidas y desconocidas, sino también «desconocidos desconocidos».[46]

Me concentro en la certeza y el riesgo, no porque sean siempre las actitudes cognitivas apropiadas, sino porque son las más comunes. Aun cuando la gente carezca de todo fundamento para tener *alguna* creencia sobre un tópico dado, a menudo siente el impulso irresistible de formarse una opinión: no una opinión específica (como en el pensamiento desiderativo), sino *la que fuere*. Hasta cierto punto, esta propensión está determinada por factores culturales. Albert Hirschman ha dicho que la mayoría de las culturas latinoamericanas «asignan un valor considerable al hecho de tener *opiniones fuertes* sobre virtualmente *todo* desde el *principio*». En esas sociedades, admitirse ignorante es admitirse derrotado. Pero, a decir verdad, la tendencia es universal. Montaigne dijo que «se da lugar a muchos engaños en el mundo, o por decirlo más osadamente, todos los engaños del mundo tienen lugar porque nos enseñan a temer el mostrar nuestra ignorancia y porque nos vemos obligados a aceptar todo cuanto no podemos refutar». La intolerancia de la incertidumbre y la ignorancia no se origina sólo en la arrogancia, sino en un deseo humano universal de encontrar significados y patrones por doquier. La mente aborrece el vacío.

Una versión particular de la tendencia a encontrar significados en el universo consiste en atribuir *agencia* a sucesos que, con igual o más verosimilitud, podrían deberse al azar. En Francia, bajo el Antiguo Régimen, la población nunca podía aceptar que la naturaleza era la única responsable de su miseria. Cuando había carestía de granos, se presumía que los acaparadores habían hecho subir los precios, aun cuando la causa real fuera una mala cosecha. A veces, la falta de grano se explicaba incluso sobre la base del deseo de las elites de hambrear al pueblo, como parte de una guerra de clases constante. Según el artículo 22 de la carta de constitución de Hamas, los judíos han estado

> detrás de la Revolución francesa, la revolución comunista y la mayoría de las revoluciones de las que hemos tenido y tenemos noticia, aquí y allá. Con su dinero formaron sociedades secretas como la francmasonería, los Rotary Clubs, los Leones y otras en diferentes partes del mundo, con el propósito de sabotear la sociedad y promover los intereses sionistas. Con su dinero pudieron controlar las naciones imperialistas e instigarlas a colonizar muchos países a fin de permitirles explotar sus recursos y difundir en ellos la corrupción.

Este molde mental conspirativo o paranoico es en gran medida inmune a la refutación, dado que los creyentes en una teoría de la conspiración la encontrarán *confirmada* por la falta de pruebas e incluso por la existencia de pruebas en contrario, que interpretan como signos de la naturaleza diabólicamente astuta de los conspiradores.

De una manera u otra, estos mecanismos generadores de errores se basan en la *motivación*. Sin embargo, el error también puede surgir de la *ignorancia*. La observación parece obvia pero, en realidad, es un poco sutil. Darwin señaló, por ejemplo, que «la ignorancia genera confianza más a menudo que el conocimiento». Sumadas, la ignorancia y la confianza son una buena receta para el error. A la inversa, cuando el círculo de luz se amplía, otro tanto hace la zona circundante de oscuridad, una razón para ser más humildes. Ciertos experimentos sugieren, de hecho, que la incompetencia no sólo causa un pobre desempeño cognitivo, sino también la incapacidad de reconocer que la propia competencia es escasa. Los incompetentes sufren una doble discapacidad.

Hay, no obstante, otra posibilidad, a saber, que «lo peligroso» sea «un *poco* de conocimiento» y no la ignorancia. Montaigne escribió que «hay ignorancia analfabeta, la cual va antes de la ciencia; y otra doctoral, que viene tras la ciencia»; Pascal dijo lo mismo más extensamente. Con la adquisición de más información, ganamos en un principio en confianza, pero a la larga ésta disminuye. Conjeturo que la hipótesis de Montaigne y Pascal es más válida que la de Darwin. No es nuestro mayor *nivel* de conocimiento sino nuestro *mayor* nivel lo que impulsa a dar a la mente un salto más allá de sí misma y generar más confianza de la que los datos justifican.

Evaluaciones subjetivas de la probabilidad

Los juicios de probabilidad pueden derivar de la observación de frecuencias objetivas o ser evaluaciones puramente subjetivas.[47] Cuando el agente está en condiciones de basarse en una gran cantidad de observaciones de situaciones similares, el método frecuentista puede redundar en buenos resultados. Si planeo hacer el mes que viene una reunión al aire libre para mi cumpleaños, y necesito formarme una opinión sobre el clima probable, lo mejor que puedo hacer es, seguramente, echar un vistazo a las estadísticas climáticas correspondientes al mismo día en

años anteriores. Pero si necesito hacerme una idea del clima de mañana, el mejor predictor es el clima de hoy. Con todo, no es el único. Los registros pasados pueden decirme si el tiempo soleado en ese día es un suceso poco común o normal. Si es poco común, el tiempo soleado de hoy pierde algo de su valor predictivo. Tal vez consulte entonces el barómetro que tengo en la pared para ver si la presión del aire está subiendo o bajando, o mire el cielo del anochecer, el vuelo de las golondrinas, etcétera.

La integración de toda esta información en un juicio de probabilidad general sobre el clima de mañana es una tarea difícil. La mayoría de la gente no es muy apta para ella. Con frecuencia, el problema no es la falta de información sino su abundancia, combinada con la carencia de un procedimiento formal para integrarla en una opinión que tome todo en cuenta. Algunas personas, sin embargo, son mejores que la mayoría cuando se trata de integrar una información vasta y difusa de diversos grados de relevancia en una evaluación global. Poseen la elusiva pero crucial cualidad del *juicio*. Los generales, empresarios y políticos exitosos suelen tenerla, y por eso su éxito. Un buen banquero central necesita tenerla; no así la mayor parte de los economistas.[48] Lomejor que podemos hacer quienes no la poseemos es reconocer esa falta y aprender a no confiar en nuestra intuición. Puedo llegar a saber, por ejemplo, que a menudo desconfío de la gente por razones que, cuando llego a entenderla, son irrelevantes. («Cuando lo conocí en quinto grado, parecía un pendenciero.») Acaso llegue, por eso, a desconfiar de mi desconfianza.[49]

Tendemos a pensar, sin embargo, que no sólo los generales, los políticos y los empresarios exitosos tienen ese juicio, sino también los expertos capacitados. En asuntos complicados de diagnóstico y pronóstico, como la identificación de los individuos psicóticos o la evaluación de la probabilidad de que una persona que solicita la libertad bajo palabra reincida en el delito, confiamos en los expertos. Debido a su experiencia, éstos son sensibles a signos reveladores que observadores sin formación podrían ignorar, o cuya significación tal vez no entenderían. Por otra parte, cuando diferentes pruebas apuntan en diferentes direcciones, los expertos pueden abrevar en su experiencia para decidir a cuál de ellas hay que atribuir mayor peso en el caso considerado. Eso es, al menos, lo que pensamos de ellos. Como casi todos nos consideramos expertos en uno u otro campo, aunque sólo sea en la predicción del comportamiento de

nuestro jefe, nuestro cónyuge o nuestros hijos, hemos invertido mucho en esta imagen de las aptitudes cognitivas superiores del experto.

Por desdicha, *esa imagen es completamente falsa*. En muchos estudios, el desempeño de los expertos en materia de diagnósticos o pronósticos se ha comparado con el desempeño de una simple fórmula mecánica basada en unas pocas variables. Esto equivale, en esencia, a comparar métodos objetivos (frecuentistas) y subjetivos. Los valores asignados a las variables se obtienen por medio de técnicas estadísticas que identifican los que tienen mayores probabilidades de predecir los resultados observados. Casi sin excepción, la fórmula funciona al menos tan bien como el experto, y muchas veces es mejor.[50] Para mencionar un solo ejemplo, en un estudio sobre el diagnóstico del disfuncionamiento cerebral progresivo basado en pruebas intelectuales, una fórmula derivada de una serie de casos y luego aplicada a una nueva muestra identificó correctamente el 83% de los nuevos casos. Grupos de clínicos experimentados e inexpertos mostraron un acierto del 63% y el 58%, respectivamente. Por otra parte, los expertos a menudo muestran fuertes discrepancias entre sí. En otro estudio, psiquiatras sumamente experimentados que observaron la misma entrevista psiquiátrica no pudieron llegar a un acuerdo sobre el diagnóstico, las motivaciones o los sentimientos del paciente. Algunos psicoterapeutas utilizan las respuestas a ambiguas manchas de tinta como indicaciones para hacer sus diagnósticos. Sin embargo, parecería ser que los pacientes son tan ambiguos para ellos como las manchas de tinta lo son para los pacientes.

Algunos errores de la inferencia estadística

No menos que los legos, los expertos se equivocan muchas veces porque ignoran principios obvios o no tan obvios del razonamiento estadístico. En un estudio, se dio a los sujetos la descripción de un joven de pelo largo que tenía la costumbre de leer poesía y se les preguntó si, a su juicio, era más probable que fuera violinista en una orquesta o camionero. La mayoría se inclinó por la alternativa del violinista, y pasó por alto de tal modo el *índice básico* de los dos grupos, es decir el número absoluto de individuos en cada uno de ellos. En el país son tanto más numerosos los camioneros que los violinistas de orquesta (y es tanta la variación entre los primeros), que, de hecho, es más probable que el joven aficionado a la poesía conduzca un camión.

Otra fuente de errores en la formación de creencias es el *sesgo de selección*. Con frecuencia, los pacientes que concurren a los centros de diálisis se muestran sorprendentemente renuentes a anotarse en la lista de espera de trasplantes de riñón. Una razón es que todos los pacientes trasplantados que ven son aquellos en quienes la operación fracasó y han debido volver a la diálisis. Montaigne hablaba de un sesgo de este tipo cuando, con referencia a Diágoras, decía que, «al mostrarle en el templo muchos exvotos y cuadros de los que habíanse salvado del naufragio, […] [le preguntaron]: "¡Y bien! Vos que pensáis que los dioses no se ocupan de las cosas humanas, ¿qué decís de tantos hombres salvados por su intercesión?" "Ocurre", respondió él, "que no están representados aquellos que se ahogaron en número muy superior"». De manera análoga, un psiquiatra que afirme que «ningún abusador de niños se detiene por decisión propia» pasa por alto el hecho de que, si alguno lo hace, es improbable que él lo conozca.

El alto mando de la fuerza aérea israelí cometió un error evidente al evaluar la eficacia relativa de las recompensas y los castigos en el entrenamiento de los pilotos. Al advertir que el desempeño de éstos mejoraba cuando se los castigaba por una mala actuación, pero no cuando se los recompensaba por una buena, llegaron a la conclusión de que el castigo era más eficiente. Ignoraban, de ese modo, el fenómeno conocido como *regresión a la media*. En cualquier serie de sucesos que están determinados en todo o en parte por el azar, hay una tendencia a que un valor extremo en una ocasión sea seguido por un valor menos extremo en la siguiente. Los hombres altos tienen hijos varones que son más bajos que ellos, y a los malos desempeños de los pilotos siguen otros que son menos malos, con prescindencia de las recompensas y los castigos. Cuando los atletas que han tenido un desempeño excepcional en una temporada no tienen un rendimiento tan bueno en la siguiente, los aficionados y los entrenadores suelen decir que el éxito los ha echado a perder, cuando lo que observamos tal vez sea simplemente una regresión a la media.

La falacia del jugador y su inversa (sin nombre) son otro ejemplo. La contratación de seguros contra terremotos aumenta de manera pronunciada después de un seísmo, pero luego sufre una caída continua a medida que el recuerdo se disipa. Tal cual hacen los jugadores que cometen el error de creer que el rojo tiene más probabilidades de volver a salir si ya ha salido varias veces seguidas, los compradores de esos seguros apelan a la

heurística de la disponibilidad para formar sus creencias. Su juicio acerca de la probabilidad de un suceso está conformado por la facilidad con que puede evocárselo, y los sucesos recientes están más al alcance que los anteriores. La caída de la emoción con el paso del tiempo (Capítulo 8) también podría ser un factor. A la inversa, las personas que viven en zonas inundables suelen creer que en el año $n + 1$ hay menos probabilidades de que se produzca una inundación si ya ha habido una en el año n. Tal cual hacen los jugadores que cometen el error de creer que el rojo tiene menos probabilidades de volver a salir si ya ha salido varias veces seguidas, esas personas forman sus creencias sobre la base de la *heurística de la representatividad*. Creen, o actúan como si creyeran, que una secuencia breve de sucesos es probablemente representativa de una secuencia más larga en la cual está inserta.

La gente capta pocas veces la relación entre los procesos aleatorios y la distribución de los resultados. Durante la Segunda Guerra Mundial, muchos londinenses estaban seguros de que los alemanes concentraban sus bombardeos de manera sistemática en ciertas zonas de la ciudad, porque las bombas caían en racimos. No entendían el principio estadístico básico de que los procesos aleatorios tienden a generar arracimamientos, y que las bombas que caen en un patrón nítido de cuadrícula cerrada serían una prueba más sólida de selección deliberada de blancos. Un hecho que nunca deja de sorprender a quienes no han tropezado nunca con él es que en un grupo de apenas veintitrés personas, la probabilidad de que dos de ellas hayan nacido el mismo día del mismo mes es de más del 50%.

Pensamiento mágico

Consideremos a continuación varias formas de *pensamiento mágico*, es decir, la tendencia a creer que uno puede ejercer una influencia causal en resultados que, en realidad, están fuera de su control. Por ejemplo, la gente hará apuestas más grandes con una moneda que todavía no ha sido lanzada que con otra que ya lo ha sido pero cuyo resultado no se ha revelado. En Proust, Robert de Saint-Loup, amigo del narrador, estaba sometido a «una suerte de creencia supersticiosa: que la fidelidad de su amante con él dependía de la suya con ella». La gente también puede pasar por alto la distinción entre la *relevancia causal* y la *relevancia diagnóstica*. En un experimento, los sujetos a quienes se había hecho creer que la cantidad de

tiempo que se podían mantener los brazos sumergidos en el agua extremadamente fría era el mejor indicador de la longevidad, los mantuvieron en ella más tiempo que quienes no habían recibido esta (falsa) información.[51] Además, al utilizar su propio comportamiento como un predictor de cómo actuarán otras personas, la gente tal vez elija la estrategia cooperativa en un dilema del prisionero, como si pudiera, de alguna manera, hacer que los otros también cooperaran. En un experimento, los sujetos cooperativos a quienes se pidió que predijeran la elección de su compañero de interacción, así como la de una persona no asociada a ellos que tenía otro compañero, se inclinaron más a predecir la cooperación de su propio camarada que la de la persona no asociada, y mostraron mayor confianza en su predicción.[52]

El calvinismo ofrece un ejemplo de este tipo de pensamiento mágico (Capítulo 3). Dada la creencia calvinista en la predestinación, parecería no haber razón para que un calvinista no se entregara a toda clase de placeres mundanos, que por hipótesis no pueden afectar su destino tras la muerte. Max Weber sostenía que, pese a ello, el calvinismo inducía a sus seguidores a adoptar un estilo de vida ascético, no para ganar la salvación, sino para conquistar la certeza subjetiva de que se contaban entre los elegidos. Podríamos interpretar que, de ese modo, Weber decía que los calvinistas confundían la relevancia causal y la relevancia diagnóstica de su comportamiento. Esto es muy explícito en una carta que los baptistas ingleses difundieron en 1770: «Toda persona que viene a Cristo para salvarse […] debe ser alentada. […] La persona que viene no debe temer no ser elegida, pues sólo los elegidos estarían dispuestos a venir». Si Dios me ha incluido entre los elegidos, también *me hará querer* determinados tipos de comportamiento.

Estos errores (y muchos otros que han sido ampliamente documentados) son en su mayor parte errores «fríos» o inmotivados, similares en algunos aspectos a las ilusiones ópticas. Otros, los errores «calientes», surgen debido a que las creencias del agente son *motivadas*, esto es, sufren una influencia indebida de sus deseos. Como veremos en el Capítulo 11, una influencia causal de los deseos sobre las creencias no es intrínsecamente irracional. Un deseo puede proporcionar una razón para invertir una cantidad específica de recursos en la adquisición de información. La información así obtenida quizá sirva como razón para sostener cierta creencia. Aunque el deseo no brinde una razón para sostenerla, interviene

en un complejo racional de formación de creencias. La cuña entre el deseo inicial y la creencia final está constituida por el hecho de que, por definición, el resultado de la búsqueda de información se desconoce en el momento de tomar la decisión de buscarla.

Formación de creencias motivadas

La influencia de los deseos sobre las creencias que acabo de mencionar es, sin discusión, compatible con la racionalidad. La apuesta de Pascal propuso una idea más controvertida. Como he explicado en el capítulo anterior, Pascal adujo que un agente que cree que hay una probabilidad no igual a cero, por pequeña que sea, de que Dios exista, debe, por la razón puramente instrumental de maximizar el valor esperado, tratar de formarse la firme creencia (del orden de la certeza) de que Dios existe, porque si existe, esa creencia le asegurará la bienaventuranza eterna. Las premisas del argumento son 1) que cierta creencia proporciona sin duda alguna la salvación, y 2) que el origen instrumental de la creencia no mengua su eficacia salvífica. Si bien ambas premisas pueden ser dudosas desde un punto de vista teológico, este asunto no debe preocuparnos aquí. La cuestión es si esa «decisión de creer» es un proyecto racional. En cierto sentido, no lo es: no puedo decidir creer a voluntad, tal como puedo decidir levantar el brazo. Podríamos, con todo, valernos de una estrategia indirecta. Al actuar *como si* creyera, argumentó Pascal, uno terminaría por creer. No obstante, el mecanismo en virtud del cual esto podría suceder es un tanto insondable.[53]

Hay otros casos en los que querríamos tal vez adoptar una creencia que creemos falsa, debido a las buenas consecuencias de sostenerla. Si quiero reducir la cantidad de alcohol que tomo pero el riesgo de convertirme en alcohólico no me motiva lo suficiente, quizá desee creer que ese riesgo es más grande de lo que ahora creo que es. En términos generales, sin embargo, no hay una tecnología confiable para adquirir esas creencias. Si el proceso no tiene un *componente de autoborrado* mediante el cual el origen de la creencia en el deseo de adoptarla se elimina de la mente consciente, es probable que ese deseo siga siendo un mero anhelo.

En el caso «no controvertido», el deseo del agente induce cierto nivel de recolección de información que, a su vez, generará alguna creencia. En el «caso controvertido», el deseo induce un comportamiento específico que a

su turno inducirá una creencia específica que el agente quiere abrigar. Ambas estrategias son indirectas. Me ocupo ahora de las creencias *directamente* formadas por la motivación. Esta formación directa puede darse de una de dos maneras, correspondientes a dos características básicas de las motivaciones: la excitación y el contenido. Así como decimos que la piedra rompe el hielo en virtud de su peso y no de su color, podemos decir que una motivación afecta una creencia no en virtud de su contenido, sino del nivel concomitante de excitación. La excitación fisiológica moderada puede mejorar la calidad de la formación de creencias, al concentrar la atención y estimular la imaginación. «Cuando un hombre sabe que van a colgarlo al cabo de quince días», dijo el doctor Johnson, «su mente se aguza maravillosamente.» Sin embargo, más allá de determinado nivel de excitación, la cognición se deteriora. En estados extremos de hambre, estrés, miedo o ansia adictiva, cuesta mucho pensar en forma ordenada porque la excitación hace difícil tener presentes los pasos previos del razonamiento. Es de presumir que la concentración mental se embota cuando falta apenas un día para la ejecución en la horca. En las pruebas de aptitud académica, una motivación muy fuerte para acertar puede llevarnos, en realidad, a equivocarnos, así como el intenso deseo de un tirador de dar en el blanco puede provocarle temblor en las manos y hacer que yerre el disparo.[54] En el capítulo siguiente sostengo que debido a su urgencia, muchas emociones pueden llevar al agente a soslayar la maquinaria normal de formación de creencias racionales. Así, es posible que la *motivación dé forma* a las creencias sin que éstas estén *motivadas*, porque el agente no tiene un deseo específico de creerlas verdaderas. La excitación *nubla* la mente, pero no la *inclina* en favor de ninguna creencia en particular.

Racionalización

Las creencias generadas por el contenido se dividen en dos variedades principales. Como ya he señalado antes, el agente puede estar motivado a abrigar *alguna creencia* sobre un tema determinado, debido a que necesita llegar al cierre de una situación o no tolera admitir su ignorancia. Otra alternativa es que esté motivado a abrigar una creencia *específica*, por ejemplo que su esposa le es fiel.[55] Los mecanismos más importantes que generan esta variedad son la racionalización, el pensamiento desiderativo y el autoengaño. La diferencia entre la primera y los dos últimos radica en la

relación con el comportamiento. En la racionalización, el comportamiento aparece en primer lugar y luego viene la creencia. (Lo cual no significa decir que las creencias, una vez adoptadas, no induzcan otros comportamientos.) En el pensamiento desiderativo y el autoengaño, observamos la secuencia contraria.

Como ejemplo de racionalización, consideremos un experimento clásico de «disonancia cognitiva». Se pide a dos grupos de sujetos que escriban una composición sobre el tema del aborto, con argumentos favorables a la posición a la que *no* adhieren (por la vida o por la libre elección). Los sujetos de un grupo reciben una suma importante por participar, mien-tras que a los del otro se les pide que lo hagan como un favor para con el experimentador. Tras escribir el texto, los integrantes del segundo grupo, no los del primero, muestran una actitud más favorable hacia la posición que han defendido. La explicación es, posiblemente, que todos los sujetos desean tener una *razón* para hacer lo que hacen. Los miembros del primer grupo pueden mencionar simplemente el dinero.[56] Los integrantes del segundo grupo pueden aludir a sus creencias (adaptadas) como razón por la cual argumentan lo que argumentan.[57]

Un proverbio francés dice: «Quien ha ofendido no puede perdonar». Si he perjudicado injustamente a una persona, quizá sea incapaz de admitirme a mí mismo que estoy en falta. Buscaré, en cambio, una falta en el otro que justifique o al menos excuse mi conducta. Los violadores dirán: «Ella se vestía provocativamente», un pretexto con el que a veces coinciden los tribunales. Quienes participen de la violencia antisemita apelarán a la historia de que como los judíos deben su éxito a medios inmorales o ilegales se hacen acreedores a un castigo. Que el comportamiento induce las creencias y no al revés es un hecho que la asombrosa flexibilidad de la racionalización muestra a las claras. Como se señaló en el Capítulo 3, la retórica del antisemitismo incluye la caracterización de los judíos como «gusanos» subhumanos, así como la idea de su omnipotencia a lo largo de la historia.

Pensamiento desiderativo

Querría ahora ocuparme del pensamiento desiderativo y el autoengaño. Estos dos fenómenos poco comprendidos tienen en común el hecho de que el deseo de que *p* sea cierto causa la creencia de que lo es. En el

pensamiento desiderativo, se trata de un proceso simple de un solo paso: el deseo es el padre del pensamiento. Más que negarse, las pruebas se ignoran. Como consecuencia, la creencia de carácter desiderativo podría ser la misma que las pruebas, de haber sido consultadas, habrían justificado.[58] Por su parte, el autoengaño, tal como suele concebírselo, implica cuatro pasos: primero, se consideran las pruebas; segundo, se forma la creencia apropiada; tercero, esta creencia es rechazada o suprimida por ser incompatible con nuestro deseo; y cuarto, el deseo provoca la formación, en su lugar, de otra creencia más aceptable. El autoengaño es un fenómeno paradójico, cuya existencia, e incluso su posibilidad, se han puesto en duda, por lo cual comenzaré con el problema más simple del pensamiento desiderativo.

Antes de sugerir un mecanismo responsable de la creación de este último tipo de pensamiento, permítaseme declarar que, a diferencia de lo que ocurre con el autoengaño, es imposible negar su existencia. Podemos negar su presencia en situaciones en las que hay mucho en juego o su influencia sobre comportamientos agregados como los de los mercados bursátiles o los electorados, pero no que ocurre. La literatura mundial bastaría por sí sola para dar testimonio de su existencia. Por otra parte, muchas creencias desiderativamente formadas actúan como premisas de la *acción* y son, por tanto, algo más que meras «cuasicreencias». Algunos fumadores que se engañan hasta creer que su hábito no es peligroso, en general o específicamente para ellos mismos, lo habrían abandonado o hubieran tratado de abandonarlo si hubiesen tenido creencias más racionales.[59] Los individuos confiados en exceso, que creen ilusoriamente ser más capaces de lo que en verdad son, pueden embarcarse en albures que en otras circunstancias habrían evitado. Las personas que se inducen a creer que son tan exitosas como otras quizá carezcan del acicate para mejorar. El siguiente es un mecanismo habitual. En primer lugar, una persona es motivada a creer que es exitosa. En segundo lugar, encuentra ciertos ámbitos de su vida en los que, en efecto, le va bien. En tercer lugar, realza la importancia de esos ámbitos para poder decirse que es exitosa en general. Por último, reduce sus esfuerzos para triunfar en otros aspectos de la existencia.

Para navegar en la vida, es instrumentalmente útil tener creencias precisas. Al mismo tiempo, las creencias pueden ser intrínsecamente agradables o desagradables, esto es, generar emociones positivas o

negativas. Si me dicen que tengo cáncer, puedo buscar un tratamiento, pero la creencia también me hará sentir muy mal. En el lenguaje de Freud, las personas regidas por el principio de realidad buscan creencias exactas, mientras que las sometidas al principio de placer buscan creencias agradables. Esta distinción sólo es válida para las creencias en sentido estricto, no para las cuasicreencias. Quienes se forjan creencias poco realistas sobre la recepción de un gran premio monetario por sus logros, pero no gastan el dinero antes de recibirlo, están a lo sumo sujetos a una forma inofensiva del principio de placer. En la variedad más nociva, su convicción de que el premio se les otorgará realmente los lleva a endeudarse.

La formación de creencias también puede tener costes. Si las creencias se forman debido a sus beneficios intrínsecos, el coste es la renuncia a los beneficios instrumentales. Dicho coste depende tanto del resultado que se producirá si la creencia motivada es falsa, en comparación con lo que ocurriría si fuera verdadera, como de la probabilidad de que sea falsa. Me referiré a estos aspectos como el *componente de resultado* y el *componente de probabilidad* de los costes. A la inversa, las creencias que se forman para proporcionar beneficios instrumentales pueden tener el coste de la renuncia a los beneficios intrínsecos. Examinaré los casos en que la formación de creencias racionales tiene costes intrínsecos de corto plazo y beneficios instrumentales de largo plazo, mientras que la creencia motivada tiene beneficios intrínsecos de corto plazo y costes instrumentales de largo plazo. Al hacer ese examen, contrastaré las explicaciones de la formación de creencias motivadas propuestas por economistas y psicólogos, para sostener que debemos basarnos en ambas.

Los economistas se concentran en los *costes*. Algunos han argumentado, por ejemplo, que los trabajadores se forman creencias motivadas sobre la seguridad laboral en función de si el beneficio de tener la creencia supera el coste. Si el beneficio psicológico de eliminar el miedo a una actividad en particular es superior al coste debido a las mayores posibilidades de accidente, el trabajador creerá que la actividad es segura. Con ello se da a entender que el agente escruta en forma inconsciente las pruebas para ver si puede permitirse adoptar la creencia promovida. Se supone, además, que no hay restricciones a la formación de creencias: el trabajador puede creer lo que quiera, sea cual fuere la información a su alcance. A mi juicio, éste es un modelo erróneo, no sólo porque no da cabida a las restricciones, sino

también porque los costes intervienen de manera equivocada. Los beneficios de las creencias motivadas se producen ahora, y los posibles costes aparecen más adelante. Para que el argumento sea eficaz, tenemos que suponer que la mente inconsciente es capaz de plantear esas soluciones de compromiso intertemporales. Como he sostenido en la introducción a esta parte del libro, no hay pruebas que respalden esa idea.

Una idea más plausible, en mi opinión, es la de que *el componente de resultados conocido de los costes* pone en funcionamiento el pensamiento desiderativo. Cuando hay poco en juego, es posible que el agente se forme una creencia sin tomar en cuenta el componente de probabilidad de los costes, es decir sin considerar las pruebas. No hay solución de compromiso, sino un proceso de dos pasos. Primero, el agente considera las apuestas en juego. Si son bajas, adopta la creencia más agradable. Si son altas, examina las pruebas y, de ser necesario, busca otras. Como es evidente, los distintos individuos entenderán de diferente manera lo que significa una «baja» apuesta. Todo lo que puede decirse es que, para un individuo en particular y en igualdad de las demás circunstancias, el pensamiento desiderativo es más probable cuando las apuestas son bajas.

Los psicólogos se concentran en las *restricciones*. Un agente que empieza a fumar tal vez sienta la tentación de hacerse la creencia ilusoria de que el hábito no es peligroso o, al menos, no lo es para él. Sin embargo, de ese modo quizás actúe limitado por sus creencias previas sobre los riesgos del fumar. La primera vez que es suspendida en un examen, una persona puede imaginarse una historia de mala suerte, pero si el mismo resultado se produce en las cuatro oportunidades siguientes, es menos probable que la historia funcione en caso de un sexto fracaso. Podemos considerar asimismo el ejemplo de las entradas caras a los espectáculos de Broadway, que presenté en el Capítulo 1. Si he pagado setenta y cinco dólares por la entrada, pero el espectáculo es muy malo, mi recuerdo de lo que he gastado será probablemente demasiado vívido para someterse a una revisión desiderativa que lo disminuya. Dada la naturaleza intangible y multidimensional de la apreciación estética, es más fácil ajustar hacia arriba mi evaluación del espectáculo. De manera similar, si bien hay pruebas de que los sucesos probables son vistos como más deseables y de que los sucesos deseables son percibidos como más probables, este último efecto tiene restricciones más sustanciales que el primero.

En un instructivo experimento, los sujetos esperaban participar en un juego sobre anécdotas de la historia con otra persona, ya fuera como compañero o como adversario. Luego de exhibírseles una muestra del desempeño de dicha persona, en la cual ésta había obtenido un puntaje perfecto, quienes esperaban tenerla de su lado (y deseaban por lo tanto que tuviera muchas aptitudes) lo juzgaron mejor en historia que quienes esperaban que fuera su adversario (y deseaban, por consiguiente, que tuviera escasas aptitudes). Al mismo tiempo, la naturaleza de la información recibida representaba una clara restricción para los sujetos, pues aun quienes esperaban que la persona en cuestión fuera su adversario la juzgaron por encima del promedio. Una limitación del experimento radica en que no brindaba a los sujetos la oportunidad de *actuar* en función de esas creencias, con las consecuencias potencialmente costosas que podrían deducirse de la subestimación de un adversario. Por lo que sabemos, podría tratarse simplemente de cuasicreencias.

En los ejemplos recién presentados, creencias fácticas previas limitan el pensamiento desiderativo. En otros casos, éste puede ser limitado por creencias causales verosímiles. Con frecuencia, el pensamiento desiderativo implica «contarse una historia»; en este caso, la idea de historia tiene estrecha relación con la idea de mecanismo de la que me ocupé en el Capítulo 2. La abundancia de mecanismos facilita encontrar alguna historia que justifique cualquier creencia que uno pretenda cierta. Tal vez yo desestime un rumor inoportuno con el proverbio «los rumores suelen mentir» y acepte uno oportuno con el proverbio «los rumores no suelen mentir». Supongamos asimismo que leo en la solicitud de inscripción a una facultad de trabajo social que la estabilidad emocional es una característica muy conveniente en las personas que se dedican a esa profesión. Si mi madre dejó de trabajar para cuidarme cuando nací, quizá yo refuerce la creencia en mi estabilidad con la historia de que la dedicación completa de los padres es beneficiosa para los hijos. En cambio, si mi madre decidió seguir trabajando y me mandó a una guardería infantil, podría llegar a decirme que los niños se benefician al estar con sus pares y tener padres que buscan su realización profesional fuera de casa.[60] Tras la derrota del equipo de fútbol de mis amores, puedo mantener mi creencia en su superioridad si la victoria del equipo rival ha sido obra de (lo que puede interpretarse como) la pura suerte. «Si la pelota no hubiese pegado en el árbitro, el ala habría recibido el pase en posición de anotar.» Si mi caballo ha terminado

en segundo lugar, puedo persistir en la creencia sobre mis aptitudes como apostador diciendo que «ha estado a punto de ganar». En un caso aún más flagrantemente irracional de pensamiento desiderativo, si aposté al 32 y salió el 33, también puedo decir que «estuve a punto de ganar», a pesar de que, de hecho, ambos números estén alejados entre sí en la ruleta.[61]

En ocasiones, sin embargo, no es fácil recurrir a una historia verosímil. Supongamos que una persona apuesta su dinero al 24. El número que sale es el 15, que está junto al 24 en la ruleta; por lo tanto, la persona confirma su creencia en sus aptitudes de apostador. Probablemente también habría considerado como confirmaciones otros resultados, por ejemplo el 5, el 10 y el 33, porque están cerca en la ruleta. Podría haber tomado de igual modo el 22, el 23, el 25 y el 26, cuyo valor numérico es más cercano, o el 20, el 21, el 26 y el 27, que son adyacentes al 24 en el paño. Así, trece de treinta y siete resultados posibles podrían estimarse confirmaciones de su habilidad de apostador. Pero eso también significa que quedan veinticuatro resultados para los cuales no es fácil apelar a una historia. Si se da alguno de ellos, incluso una persona propensa al pensamiento desiderativo y muy motivada a adoptar una creencia específica podría tener que hacer frente a los hechos.

Autoengaño

Ocupémonos ahora del espinoso problema del autoengaño. En la vida cotidiana, la gente parece engañarse con cosas como su peso, su salud, sus hábitos de bebedor, su tendencia a postergar o la fidelidad de sus cónyuges. En una situación típica, la persona recibe información que sugiere que algo anda mal, pero luego omite hacer más averiguaciones para llegar a una conclusión definitiva. Al mirarme al espejo, advierto que peso más de lo debido, pero es difícil decir cuánto. Si evito pesarme, puedo decirme que probablemente sean apenas unos kilos y que soy capaz de rebajarlos cuando quiera. Una mujer siente un bulto en uno de sus pechos pero no concierta una cita con su médico para determinar si es benigno o maligno. En tales casos, la falta de un conocimiento preciso facilita el autoengaño. La mujer no llega a la conclusión de que tiene cáncer y luego elimina la creencia. Antes bien, sospecha que *podría* tenerlo.

Los siguientes rasgos caracterizan este caso, que trataré como paradigmático:

1. La sospecha inicial de cáncer toma la forma de una creencia de baja probabilidad.
2. Está acompañada de una firme creencia de que, si de veras tiene cáncer y no hace nada al respecto, el resultado será casi ineluctablemente fatal.
3. También la acompaña la firme creencia de que, si tiene cáncer y hace algo al respecto, el resultado, pese a ello, puede llegar a ser fatal y, aunque no lo sea, el tratamiento será muy desagradable.
4. La mujer no se pregunta en ningún momento si el dolor del tratamiento (comparado con el dolor de una enfermedad no tratada) compensa el riesgo diferencial de mortalidad.
5. En cambio, se limita a abstenerse de consultar al médico para averiguar si tiene cáncer.

En este ejemplo, los rasgos decisivos son 1) y 3). Como la creencia inicial es de baja probabilidad, los costes de modificarla son pequeños. La mujer puede concentrarse con facilidad en muchas historias que habrá oído acerca de bultos inofensivos y alarmas innecesarias. *No obstante, en ausencia del rasgo 3), no tiene motivaciones para dar nueva forma a sus creencias*. Si supiera que hay un tratamiento gratuito e indoloro y de eficacia comprobada, no tendría motivación para no ir al médico. No hay forma conocida de irracionalidad que favorezca la tendencia a bloquear la evitación *sin coste* de desastres de baja probabilidad, convirtiendo por ejemplo una baja probabilidad subjetiva en una probabilidad cero.

Así concebido, el autoengaño no implica el mantenimiento simultáneo de dos creencias contradictorias, una sostenida en forma consciente y otra, en forma inconsciente. Cuando la evaluación inicial de la probabilidad es reemplazada por otra, la primera *desaparece para siempre* en vez de quedar relegada al inconsciente. Muchos autores consideran que esta contradicción es la característica central del autoengaño. No es obligatorio, sin embargo, que así sea. La persona que se autoengaña no es (o no es necesariamente) como la persona que odia ver gatos pero comprueba que, para apartar la vista de ellos, antes debe advertir su presencia. Se asemeja más a quien ve una sombra en la oscuridad que podría ser un gato, pero también cualquier otra cosa. Para satisfacer su aversión a los gatos, en un principio reinterpreta la sombra y luego se abstiene de acercarse para constatar si se trata, en efecto, de ese animal.

Sospecho que la psicología o la filosofía no pueden hacer mucho más que la literatura para dar vida al autoengaño. En *Por el camino de Swann*, Proust se refiere a las reflexiones de Swann sobre «aquella época en que le hablaban de Odette [su amante] como de una "mujer mantenida"», y lo incongruente que le parecía esa descripción comparada con la Odette que él conocía. Una cadena de asociaciones lo lleva a pensar en su banquero y le

recuerda que debe retirar algo de dinero para ayudarla a salir de algunas dificultades materiales.

Y entonces se preguntó de pronto si aquello que estaba haciendo no era cabalmente «mantenerla» […], y si no se podía aplicar a Odette […] ese dictado que tan incompatible con ella se figuraba Swann de «mujer mantenida». Pero no pudo ahondar en esa idea, porque un acceso de pereza de espíritu, que en él eran congénitos, intermitentes y providenciales, llegó en aquel momento y apagó todas las luminarias de su inteligencia, tan bruscamente como, andando el tiempo, cuando hubiera luz eléctrica, podría dejarse una casa a oscuras en un momento. Su pensamiento anduvo a tientas un instante por las tinieblas; se quitó los lentes, limpió sus cristales, se pasó las manos por los ojos, y no volvió a vislumbrar la luz hasta que tuvo delante una idea completamente distinta, a saber: que el mes próximo convendría mandar a Odette seis mil o siete mil francos en vez de cinco mil, por la sorpresa y la alegría que con eso iba a darle.

* * *

NOTA BIBLIOGRÁFICA

Evidencias de muchos de los descubrimientos aquí mencionados podrán encontrarse en los siguientes libros de referencia: Daniel Kahneman, Paul Slovic y Amos Tversky (comps.), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge y Nueva York, Cambridge University Press, 1982; David Bell, Howard Raiffa y Amos Tversky (comps.), *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, Cambridge y Nueva York, Cambridge University Press, 1988; Terry Connolly, Hal Arkes y Kenneth R. Hammond (comps.), *Judgment and Decision Making: An Interdisciplinary Reader*, Cambridge y Nueva York, Cambridge University Press, 2000; Daniel Kahneman y Amos Tversky (comps.), *Choices, Values, and Frames*, Cambridge y Nueva York, Cambridge University Press, 2000; Thomas Gilovich, Dale Griffin y Daniel Kahneman (comps.), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge y Nueva York, Cambridge University Press, 2002; Colin Camerer, George Loewenstein y Matthew Rabin (comps.), *Advances in Behavioral Economics*, Nueva York, Russell Sage, 2004, e Isabelle Brocas y Juan Carrillo (comps.), *The Psychology of Economic Decisions*, dos volúmenes,

Oxford y Nueva York, Oxford University Press, 2003-2004. La mayor tendencia a atribuir cooperación a asociados que a no asociados está documentada en Lawrence Messé y John Sivacek, «Predictions of others' responses in a mixed-motive game: self-justification of false consensus?», *Journal of Personality and Social Psychology*, 37, 1979, págs. 602-607. La doble incompetencia del ignorante se documenta en Justin Kruger y David Dunning, «Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments», *Journal of Personality and Social Psychology*, 77(6), 1999, págs. 1121-1134. Se encontrará un refinado estudio de la falta de confiabilidad de los juicios de algunos expertos («puercoespines») y los juicios un tanto más confiables de otros («zorros») en Philip Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?*, Princeton, Princeton University Press, 2005. Los datos sobre seísmos e inundaciones están en Paul Slovic, *The Perception of Risk*, Sterling (Virginia), Earthscan, 2000. Sobre la (i)lógica de las teorías conspirativas, véase Brian Keeley, «Of conspiracy theories», *Journal of Philosophy*, 96(3), 1999, págs. 109-126. Sobre las teorías de la hambruna, véase Steven Kaplan, «The famine plot persuasion in eighteenth-century France», *Transactions of the American Philosophical Society*, 72, tercera parte, 1982, y François Ploux, *De bouche à oreille: naissance et propagation des rumeurs dans la France du XIX^e siècle*, París, Aubier, 2003. Richard Hofstadter, *The Paranoid Style in American Politics, and Other Essays*, Cambridge (Massachusetts), Harvard University Press, 1964, es un estudio del pensamiento conspirativo. Su papel en Medio Oriente es el tema de Daniel Pipes, *The Hidden Hand: Middle East Fears of Conspiracy*, Nueva York, St. Martin's Press, 1996. Mis comentarios sobre el tratamiento que dos economistas (George Akerlof y Matthew Rabin) y una psicóloga (Ziva Kunda) dan a la formación de creencias motivadas tienen mayor elaboración en «Costs and constraints in the economy of mind», en Isabelle Brocas y Juan Carrillo (comps.), *The Psychology of Economic Decisions*, vol. 2, Oxford y Nueva York, Oxford University Press, 2004, págs. 3-14. El mejor panorama ge-El mejor panorama general del autoengaño se encontrará en un número especial de *Behavioral and Brain Sciences*, 20(1), 1997, organizado alrededor de un artículo de Alfred Mele, «Real self-deception», págs. 91-136.

Notas al pie

[43] Tras la caída del comunismo, una mujer de la ex Alemania Oriental dijo en un mitin público que su generación había sido educada desde la niñez para acatar, mantenerse en fila. Una esquizofrenia de largo plazo los había *vaciado* como personas. De modo que, siguió diciendo esta mujer, ahora ella no podía de buenas a primeras «hablar abiertamente» ni «decir lo que pensaba». En realidad, ni siquiera sabía con precisión qué pensaba.

[44] Las cosas son un poco más complicadas. Cuando la gente se ve frente a probabilidades numéricas muy altas, como un 99,9%, es posible que se pongan en funcionamiento dos mecanismos (en el sentido que se les ha dado en el Capítulo 2): la diferencia entre incertidumbre y alta probabilidad se pasa por alto o se exagera. (De manera similar, también las probabilidades muy bajas se ignoran o se exageran.) Esta complicación no surge, sin embargo, en el ejemplo del efecto de certidumbre dado en el texto.

[45] Tal vez puedan asignar probabilidades ordinales, esto es, decir que un resultado tiene más probabilidades de ocurrir que otro, sin poder decir cuánto más probable es.

[46] También son posibles casos más complejos. Yo podría acaso asignar probabilidades a algunos resultados, y no ser siquiera capaz de especificar otros.

[47] En un análisis más profundo, el primer método (objetivo) se reduce al segundo (subjetivo), dado que, para ser útiles, los datos objetivos siempre necesitan una interpretación subjetiva. No obstante, a mis efectos prácticos la distinción es clara y útil.

[48] Con referencia a Alan Greenspan, Paul Krugman señaló en un artículo de *The New York Times* (28 de octubre de 2005) que, si bien desconfiaba de los modelos formales, tenía «la habilidad de adivinar, a partir de datos fragmentarios y a veces contradictorios, de dónde soplaba el viento de la economía».

[49] Saber que uno puede ser parcial es una cosa; otra es ser capaz de corregirlo. Los estudios muestran que los intentos de eliminar la parcialidad de nuestros juicios son de escaso valor, dado que caemos con facilidad en las trampas de la corrección insuficiente, la corrección innecesaria o la hipercorrección. Quizás aprendamos a desconfiar de nuestro juicio, pero es más arduo mejorarlo. Si fuéramos capaces de hacerlo, tal vez no necesitaríamos hacerlo.

[50] ¡Esta superioridad se mantiene aun cuando nos limitemos a asignar iguales valores a todas las variables!

[51] Esta distinción entre causa y síntoma no siempre es evidente. Aún en 1959, el gran estadístico R. A. Fisher, suponiendo un rasgo genético que predisponía al individuo tanto a fumar como al cáncer, sostenía que el hábito de fumar era diagnóstico del cáncer y no su causa. (Aunque es cierto que, en la época, Fisher estaba en la nómina de pagos de las compañías tabacaleras). Considérese asimismo el descubrimiento, examinado en el Capítulo 2, de que, cuanto más tiempo haya pasado un individuo desempleado, menos probabilidades tendrá de encontrar trabajo en un plazo determinado. La duración del paro podría ser simplemente diagnóstica de la empleabilidad, o hacer un aporte causal (en virtud de la desmoralización y otros factores) a las oportunidades de encontrar empleo.

[52] Esta discrepancia nos permite excluir la posibilidad de que la atribución de un comportamiento cooperativo al compañero de interacción se haya debido meramente al «efecto de falso consenso» (Capítulo).

[53] Por razones que se explicarán más adelante, la reducción de la disonancia no es un mecanismo plausible.

[54] Y de nada sirve decirnos o que nos digan que es preciso relajarse, dado que el estado de relajamiento, como el estado de sueño, es esencialmente un subproducto.

[55] O, como en el caso de Otelo, infiel. Esas «creencias contramotivadas» se han mencionado brevemente en el Capítulo 2 y serán objeto de otro examen en el Capítulo 23.

[56] El castigo por *no* participar también sería una razón suficiente. Esto explica por qué los ciudadanos de los regímenes comunistas podían tener un sistema de doble registro sin que su entusiasmo externo debilitara su rechazo interno.

[57] En la apuesta de Pascal, la razón para actuar como si uno creyera es tan abrumadoramente fuerte (la perspectiva de la bienaventuranza eterna) que el creyente no necesita buscar otra explicación de su comportamiento.

[58] La ignorancia de este aspecto podría ser el origen de la formación de creencias irracionales. Como a menudo es fácil detectar el funcionamiento de la formación de creencias motivadas en otros, tendemos a descreer de la conclusión alcanzada de este modo, sin detenernos a ver si las pruebas pueden, de hecho, justificarla. Hasta 1990, más o menos, yo creía, junto con la mayoría de mis amigos, que en una escala del mal de cero a diez (el peor), el comunismo tenía entre siete y ocho puntos. Las recientes revelaciones me hacen creer que el número apropiado es diez. La razón de mi errónea percepción de las pruebas no era una creencia idealista en el sentido de que el comunismo era un ideal valioso que había sido traicionado por los comunistas de carne y hueso. De haber sido así, yo habría sido una mera víctima del pensamiento desiderativo o el autoengaño. Lo que me descaminaba era, antes bien, el carácter histérico de quienes habían afirmado todo el tiempo que el puntaje adecuado era un diez. Mi ignorancia de sus afirmaciones no era del todo irracional. En promedio, es sensato desechar las manifestaciones de quienes son notoriamente histéricos. Sin embargo, aun los histéricos pueden acertar, si bien por las razones equivocadas. Como yo sentía y todavía creo que muchos de estos feroces anticomunistas habrían dicho lo mismo cualesquiera fuesen las pruebas, no podía creer que lo que decían correspondía a la evidencia concreta. Cometía el error de verlos como el reloj que siempre está una hora atrasado, y no como el reloj descompuesto que muestra la hora correcta dos veces por día.

[59] Como en el caso del alcohol, el abandono del hábito tal vez exija la creencia irracional de que el fumar es *más* peligroso de lo que efectivamente es.

[60] En rigor de verdad, no se encuentran diferencias consistentes en el desarrollo ulterior de los niños criados en uno u otro entorno.

[61] Al mismo tiempo, si el número está cerca del ganador, la decepción puede ser más grande. Algunas loterías nacionales ofrecen pequeños «premios consuelo» a quienes «han estado a punto de ganar».

Capítulo 8

Emociones

El papel de las emociones

Las emociones intervienen de tres maneras en la vida humana. En su máximo nivel de intensidad, son las *fuentes más importantes de felicidad y desdicha*, muy por encima de los placeres hedonistas y el dolor físico. El radiante amor de Anne Elliot al final de *Persuasión* es una felicidad insuperable. A la inversa, la emoción de la vergüenza puede ser absolutamente devastadora. Voltaire escribió: «Ser objeto de desprecio para aquellos con quienes uno vive es algo que nadie ha sido y nunca será capaz de soportar».

La vergüenza también ilustra la segunda manera de intervención de las emociones, a saber, en su *influencia sobre el comportamiento*. En el Capítulo 4 he mencionado varios casos de personas que se suicidaron debido a la abrumadora emoción de la vergüenza. En este capítulo examinaré sobre todo las *tendencias a la acción* que se asocian a las emociones. La traducción de esas tendencias a un comportamiento real nos ocupará en capítulos posteriores.

En tercer lugar, las emociones intervienen en razón de su influencia sobre *otros estados mentales*, especialmente las creencias. Cuando un deseo de que exista cierto estado cuenta con el respaldo de una emoción fuerte, la tendencia a creer que existe puede ser irresistible. Como Stendhal dice en *Del amor*: «Desde el momento en que ama, el hombre más sensato no ve ya nada *tal como es*. […] No atribuye nada al azar; pierde el sentido de la

probabilidad; una cosa imaginada es una cosa existente por el efecto sobre su felicidad». En *En busca del tiempo perdido*, Proust se ocupa del mismo tema a lo largo de cientos de páginas, con más variaciones y giros de los que se habrían creído posibles.

¿Qué son las emociones?

Antes de considerar cada uno de esos aspectos con mayor detalle, es preciso que diga *qué son* y *cuáles son* las emociones. No hay una definición convenida de lo que debe entenderse por emoción: esto es, no hay una lista aceptada de condiciones suficientes y necesarias. Ni siquiera hay una lista aceptada de condiciones necesarias. Si bien examinaré una gran cantidad de rasgos comunes de los estados que, con anterioridad a todo análisis, entendemos como emociones, todos ellos tienen contraejemplos. Es decir que en el caso de cualquiera de esos rasgos, hay algunas emociones o incidencias emocionales en las que está ausente. Tal vez estimemos que las tendencias a la acción son cruciales para la emoción, pero las emociones estéticas nos ofrecen un contraejemplo. Quizá consideremos que una «vida media breve», esto es, la tendencia a decaer rápidamente, es una característica esencial de la emoción, pero en algunos casos el amor romántico no correspondido (como el de Cyrano de Bergerac) o el deseo apasionado de venganza pueden persistir durante años. Acaso pensemos que las creencias desencadenan las emociones, pero ¿cómo explicamos entonces que la gente pueda sufrir perturbaciones emocionales al leer relatos o mirar películas que son claramente ficticios? Podríamos dar muchos otros ejemplos de presuntos rasgos universales que en algunos casos resultan estar ausentes.[62]

A la luz de este problema, la respuesta natural consiste en negar que la «emoción» sea una categoría científica útil. En el lenguaje de los filósofos, las emociones no parecen constituir una *clase natural*. A despecho de su diferencia, ballenas y murciélagos, en cuanto mamíferos, pertenecen a la misma clase natural; no así las ballenas y los tiburones, a pesar de su similitud, y tampoco los murciélagos y los pájaros. La ira y el amor tienen en común la capacidad de nublar y predisponer la mente, pero esta similitud no los convierte en una clase natural. Para ver hasta dónde puede errar el camino este razonamiento por analogía, podemos señalar que la ingesta de anfetaminas y el amor romántico producen muchos efectos semejantes:

agudeza de la conciencia, mayor energía, menor necesidad de sueño y comida y sentimientos de euforia. Sin embargo, nadie afirmaría, supongo, que los dos estados corresponden a la misma clase natural.[63]

Cuando se trata de la explicación de las ciencias sociales, este acertijo puede quedar sin resolver. Podemos concentrarnos en las apariciones de emociones en las cuales se observa con regularidad una cantidad determinada de rasgos y preguntarnos cómo pueden éstos ayudarnos a explicar el comportamiento u otros estados mentales. El hecho de que en otras instancias que intuitivamente se consideran emociones algunos de esos rasgos estén ausentes es interesante desde un punto de vista conceptual, pero no disminuye su eficacia explicativa en los casos en que están presentes. Los rasgos en los que quiero hacer hincapié son los siguientes:

- *Antecedentes cognitivos*. Las emociones son desencadenadas por creencias, a menudo una nueva que el agente ha adoptado. También pueden tener otras condiciones causales (nos irritamos con más facilidad cuando estamos cansados), pero la presencia de éstas no provocará por sí misma la aparición de la emoción, así como un camino resbaladizo no es por sí solo la causa de un accidente automovilístico.
- *Excitación fisiológica*. Junto con las emociones se producen cambios en las pulsaciones cardíacas, la conductancia eléctrica de la piel, la temperatura corporal, la presión sanguínea, la respiración y muchas otras variables.
- *Expresiones fisiológicas*. Junto con las emociones aparecen signos observables característicos, como la postura corporal, el timbre de la voz, rubores y enrojecimientos (por la vergüenza), sonrisas o dientes exhibidos en son de amenaza, risas y ceños fruncidos, llantos y lamentos y cólera blanca o roja (según se manifiesta en la palidez y el encendimiento, respectivamente).
- *Tendencias a la acción*. Las emociones se acompañan de tendencias o impulsos a la ejecución de acciones específicas. Aunque estas tendencias pueden no desembocar en un comportamiento real, son algo más que disposiciones: formas incipientes de comportamiento y no una mera conducta potencial.
- *Objetos intencionales*. A diferencia de otros fenómenos viscerales como el dolor o el hambre, las emociones *se refieren* a algo. Pueden tener «objetos proposicionales» («Me indigna que…») o no proposicionales («Estoy indignado con…»).
- *Valencia*. Éste es un término técnico para aludir a la dimensión dolorosa y placentera de las emociones tal como las experimentamos. Como se ha señalado, la valencia podría oscilar entre la dicha resplandeciente de Anne Elliot y la vergüenza aplastante de los consumidores de material pedofílico expuestos a la luz del día.

¿Las emociones no tienen también, como los colores, *sensaciones* cualitativas específicas? La vergüenza y la culpa, por ejemplo, parecen despertar *sensaciones* diferentes, de una manera que no puede reducirse al hecho de que la primera es más intensamente displacentera. Hay pruebas de que podría insertarse un electrodo en el cerebro de una persona y hacerla sentirse triste, avergonzada o temerosa, aun cuando no fuera capaz de

identificar una causa o un objeto de la sensación. Por importante que este aspecto pueda resultar para nuestra comprensión de la emoción, todavía no se lo entiende lo bastante para sugerir hipótesis causales específicas.

¿Qué emociones hay?

Enumeraré y describiré brevemente alrededor de dos docenas de emociones, sin pretender que esta clasificación sea superior a muchas otras que se han propuesto. Mi objetivo es dar cierta idea de las emociones que tienen importancia intrínseca o causal en la vida social, no procurar satisfacer las (legítimas) inquietudes de los teóricos de la emoción. En particular, no tendré nada que decir en lo concerniente a la identificación de las emociones «básicas» y «no básicas».

Un grupo importante está constituido por las *emociones evaluativas*. Éstas implican una evaluación positiva o negativa del comportamiento o el carácter de uno mismo o de algún otro.[64] Si el comportamiento de otra persona desencadena una emoción, ese comportamiento puede estar dirigido hacia uno mismo o hacia un tercero. Como resultado de estas distinciones, podemos mencionar en total diez (u once) emociones:

- La *vergüenza* es provocada por una creencia negativa sobre el carácter de uno mismo.
- El *desprecio* y el *odio* son provocados por creencias negativas acerca del carácter de otro. La idea de que el otro es inferior induce el desprecio, mientras que la idea de que es malo induce el odio.
- La *culpa* es provocada por una creencia negativa sobre la acción de uno mismo.
- La *ira* es provocada por una creencia negativa sobre la acción de otro hacia uno mismo.
- La *indignación cartesiana*[65] es provocada por una creencia negativa acerca de la acción de otro hacia un tercero.
- La *altivez* es provocada por una creencia positiva sobre el carácter de uno mismo.
- El *agrado* es provocado por una creencia positiva sobre el carácter de otro.
- El *orgullo* es provocado por una creencia positiva sobre la acción de uno mismo.
- La *gratitud* es provocada por una creencia positiva sobre la acción de otro hacia uno mismo.
- La *admiración* es provocada por una creencia positiva acerca de la acción de otro hacia un tercero.

Segundo, hay un conjunto de emociones generadas por la idea de que otra persona está en posesión, merecida o inmerecida, de un bien o un mal.[66] El blanco de estas emociones no es la acción individual ni el carácter individual, sino un estado de cosas. De conformidad con el análisis que hace Aristóteles en la *Retórica*, podemos distinguir seis (o siete) casos:

- La *envidia* es causada por el bien merecido de otra persona.
- La *indignación aristotélica* es causada por el bien inmerecido de otra persona.[67] La emoción íntimamente relacionada del *resentimiento* es causada por la inversión de una jerarquía de prestigio, cuando un grupo o un individuo antes inferiores alcanzan una posición dominante.
- La *simpatía* es causada por el bien merecido de otra persona.
- La *compasión* es causada por el mal inmerecido de otra persona.
- El *resentimiento* es causado por el mal inmerecido de otra persona.
- El *regodeo* es causado por el mal merecido de otra persona.

Tercero, hay emociones positivas o negativas generadas por la idea de cosas buenas o malas que nos han ocurrido o nos ocurrirán: *alegría* y *pena*, con sus distintas variedades y análogos. Como muchos han señalado, los malos sucesos del pasado también pueden generar emociones positivas en el presente, y los buenos sucesos, emociones negativas. Así, en la principal colección de dichos proverbiales de la Antigüedad, las *Sentencias* de Publio Siro, encontramos tanto «la rememoración de peligros pasados es agradable» como «la pasada felicidad aumenta la presente desdicha».

Todas las emociones mencionadas hasta aquí son inducidas por creencias que se sostienen (o pueden sostenerse) con certeza. También hay emociones (la *esperanza*, el *miedo*, el *amor* y los *celos*) que implican en esencia creencias sostenidas al modo de la probabilidad o la posibilidad. Estas emociones son generadas por la idea de cosas buenas o malas que pueden o no ocurrir en el futuro, y de buenos o malos estados de cosas que pueden o no existir en el presente.[68] En general, exigen que el suceso o estado en cuestión se considere como algo más que meramente concebible; esto es, debe haber una posibilidad no desdeñable o una «historia causal fácil» que pueda ocurrir o existir realmente. La idea de ganar el gordo en la lotería puede suscitar esperanza, pero no la «difícil» idea de recibir un gran regalo de un millonario desconocido. Estas emociones también parecen requerir que el suceso o estado no llegue a concebirse como seguro. Si *sé* que estoy a punto de ser ejecutado, tal vez sienta más desesperación que miedo. Según Stendhal, el amor se marchita tanto cuando uno está seguro de ser correspondido como cuando uno está seguro de no serlo. Según La Rochefoucauld, los celos pueden desaparecer en el momento mismo en que uno *sabe* que la persona amada está enamorada de otro.

Algunas emociones son generadas por pensamientos *contrafácticos* sobre lo que podría haber ocurrido o lo que uno podría haber hecho. La *desilusión* es la emoción que se produce cuando un suceso positivo esperado no se materializa.[69] La *pesadumbre* es la emoción que aparece

cuando nos damos cuenta de que, de haber tomado una decisión diferente, podríamos haber hecho realidad un suceso positivo esperado. Las contrapartidas positivas de estas emociones (causadas por la no aparición de sucesos negativos) se denominan a veces *júbilo* y *regocijo*, respectivamente. (En el lenguaje cotidiano, ambas suelen confundirse bajo el encabezado de *alivio*.) Mientras que la desilusión y el júbilo implican comparaciones de diferentes resultados generados por diferentes estados del mundo para una decisión determinada, la pesadumbre y el regocijo implican comparaciones causadas por diferentes decisiones dentro de un único estado. En algunos casos, los sucesos negativos pueden atribuirse a cualquiera de los dos orígenes. Si la lluvia me empapa en el camino al trabajo, puedo adjudicar el percance a un suceso meteorológico azaroso o al hecho de no haber llevado un paraguas. Aunque quizá prefiera la primera formulación, esta instancia de pensamiento desiderativo podría estar sujeta a limitaciones de la realidad (Capítulo 7) si se da el caso de que me he enterado de un pronóstico de lluvia justo antes de salir de mi casa.

Las emociones y la felicidad

El papel de las emociones en la generación de la felicidad (o la desdicha) sugiere la idea de un «producto nacional bruto de felicidad». Las mediciones habituales del desempeño económico son, desde luego, más objetivas. No obstante, lo que nos preocupa no es en última instancia la objetividad en el sentido de la mensurabilidad física. La razón por la que queremos conocer la producción económica es que contribuye al bienestar o la felicidad *subjetiva*. Por otra parte, la felicidad puede surgir de fuentes que no se prestan a tipo alguno de medición cuantitativa objetiva. En 1994, cuando Noruega fue la anfitriona de los Juegos Olímpicos de Invierno, el país tuvo que construir nuevas instalaciones para las competencias y el alojamiento de los participantes, a un coste considerable. En la columna de los ingresos podríamos incluir el dinero gastado por los visitantes al país y los espectadores de las competencias, así como las rentas generadas por las nuevas construcciones en el futuro. Los economistas que efectuaron estos cálculos no creían que el saldo de los juegos fuera equilibrado. Yo estoy absolutamente seguro, sin embargo (aunque, por supuesto, no puedo probarlo), de que si incluyéramos los beneficios emocionales para la población noruega, los juegos mostrarían un enorme superávit. El número

inesperadamente alto de medallistas dorados noruegos suscitó un humor de euforia colectiva, que fue mucho más grande *porque* las victorias eran, precisamente, inesperadas. La cantidad «objetiva» de triunfos debió mucho de su impacto al elemento de la sorpresa subjetiva.[70] Más recientemente, la victoria del equipo francés de fútbol en la Copa del Mundo de 1998, así como su derrota en 2002, generaron sentimientos de euforia y desazón que debían gran parte de su intensidad al hecho de la sorpresa.

En general, es difícil comparar los componentes emocionales del bienestar o la prosperidad con otros componentes. El hecho de que las emociones positivas, en sus niveles máximos de intensidad, contribuyan más a la felicidad que el simple bienestar hedonista no prueba nada, a menos que sepamos con cuánta frecuencia se producen los episodios intensos. Además, no entendemos si, y en qué medida, la propensión a los picos emocionales va a la par con la propensión a las caídas emocionales. Si lo hace, ¿una vida de contento uniforme es en general más feliz que otra que alterne entre la euforia y la disforia? Como señaló Montaigne, la respuesta depende de las oportunidades brindadas por el entorno.

> Y si me dicen que la ventaja de tener la sensibilidad fría y roma para los dolores y los males, trae consigo el inconveniente de hacernos también, por consiguiente, menos agudos y sensitivos para gozar de los bienes y de los placeres, es verdad; mas la miseria de nuestra condición hace que no tengamos tantas ocasiones para gozar como para huir.

El ideal de la extinción de las emociones que encontramos en muchas filosofías antiguas, particularmente el estoicismo y el budismo, surgió en sociedades donde el entorno tal vez brindara más oportunidades a las emociones de valencia negativa. Puesto que escribía durante las guerras de religión que devastaban Francia, es muy posible que Montaigne se encontrara en la misma situación.

La emoción y la acción

El vínculo mediador entre la emoción y la acción es el de una tendencia a la acción (o la prontitud para emprenderla). También podemos pensar en una tendencia a la acción como una preferencia temporaria. Cada una de las grandes emociones parece haber asociado a ella una (o algunas) de esas tendencias (véase el Cuadro 8.1).

Aunque la ira y la indignación cartesiana inducen la misma tendencia a la acción, la de la primera es más fuerte. Los experimentos muestran que los sujetos están dispuestos a pagar un mayor coste para perjudicar a quien los ha perjudicado que para perjudicar a quien ha perjudicado a un tercero. Finalizada la Segunda Guerra Mundial, los estadounidenses solían mostrar mayor afán en castigar a los nazis que habían infligido malos tratos a prisioneros de guerra estadounidenses que a quienes eran responsables del Holocausto. Una excepción, que confirma el principio, eran los miembros judíos de los Gobiernos de Roosevelt y Truman.

Emoción	*Tendencia a la acción*
Ira o indignación cartesiana	Hacer sufrir al objeto de la emoción
Odio	Hacer que el objeto del odio deje de existir
Desprecio	Aislar; evitar
Vergüenza	«Esconderse bajo tierra»; escapar; suicidarse
Culpa	Confesar; enmendarse; herirse
Envidia	Destruir el objeto envidiado o a su poseedor
Miedo	Huir; pelear
Amor	Acercarse y tocar al otro; ayudar al otro; complacer al otro
Compasión	Consolar o aliviar la aflicción del otro
Gratitud	Ayudar al otro

CUADRO 8.1

Las emociones de la ira, la culpa, el desprecio y la vergüenza tienen íntimas relaciones con las normas morales y sociales. Los transgresores de normas pueden sufrir culpa o vergüenza, mientras que quienes observan la transgresión sienten ira o desprecio. Las estructuras de estas relaciones difieren tal como se muestra en la Figura 8.1.

Las normas sociales, examinadas con más detenimiento en el Capítulo 22, están mediadas por la exposición a los otros. Por eso los suicidios mencionados al comienzo de este capítulo sólo se produjeron cuando las acciones vergonzosas llegaron a conocimiento del público. Como he sostenido en el Capítulo 5, las normas morales difieren en este aspecto.[71]

Algunas de las tendencias a la acción parecen apuntar a «restablecer el equilibrio moral del universo». El hecho de perjudicar a quienes nos

perjudican y ayudar a quienes nos ayudan es, al parecer, una manera de *saldar las cuentas*. Esto puede ser cierto en algunos casos. No obstante, la teoría de la perspectiva (Capítulo 12) da a entender que «dos ojos por uno» puede ser una mejor descripción de la tendencia a la acción de la ira que «ojo por ojo».[72] Aunque muchos actos de gratitud pueden realizarse simplemente para liberarse de una deuda, es posible que otros surjan de un auténtico sentimiento de buena voluntad hacia nuestro benefactor. El punto de vista del equilibrio moral tiene más peso en el caso de la culpa, en la que la tendencia a la acción consistente en enmendarse es de un carácter explícitamente restaurativo. Por otra parte, cuando el agente no puede deshacer el daño que ha hecho, puede restablecer el equilibrio dañándose a sí mismo en igual medida. Si he hecho trampas en el impuesto sobre la renta y descubro que la agencia impositiva del Estado no acepta un pago anónimo por la suma que debo, puedo restablecer el equilibrio quemando el dinero.

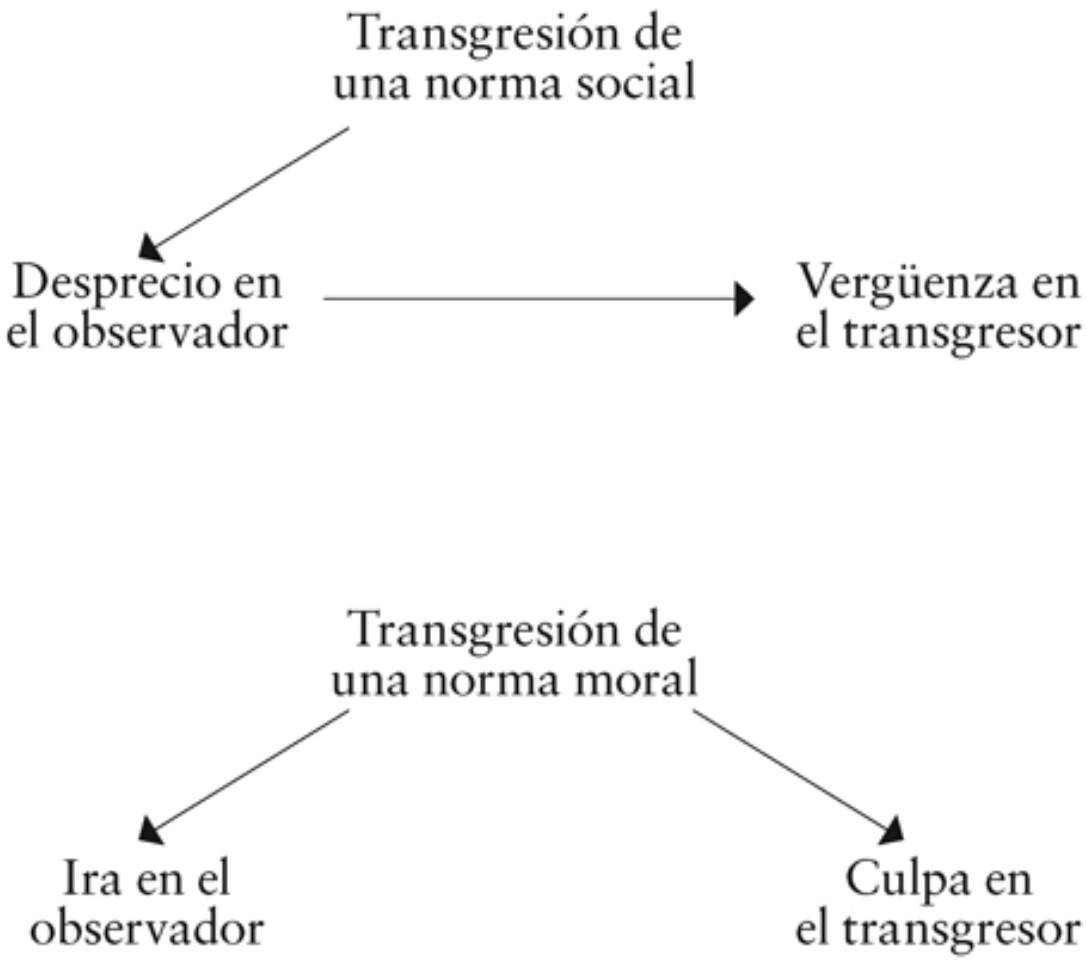


FIGURA 8.1

Las tendencias a la acción emocional no se limitan a inducir un deseo de actuar. También inducen un deseo de *actuar lo antes posible*. Para situar esta idea en contexto, querría distinguir entre *impaciencia* y *urgencia*. Defino la impaciencia como la preferencia por una recompensa temprana sobre una recompensa ulterior, es decir, cierto grado de descuento temporal. Como he señalado en el Capítulo 6, las emociones pueden hacer que un agente atribuya menos importancia a las consecuencias temporal-mente remotas de la acción presente. Defino la urgencia, otro efecto de la emoción, como una preferencia por la acción temprana sobre la acción ulterior. La distinción tiene su ilustración en el Cuadro 8.2.

En cada caso, el agente puede realizar una y sólo una de dos acciones, A o B. En el caso 1, estas opciones están a su alcance al mismo tiempo; en los

casos 2 y 3, en tiempos sucesivos. En el caso 2, las recompensas (cuya magnitud indican los números) aparecen en el mismo tiempo ulterior; en los casos 1 y 3, en tiempos ulteriores sucesivos. Supongamos que en un estado no emocional, el agente escoge B en todos los casos, pero que en un estado emocional escoge A. En el caso 1, la elección de A se debe a una impaciencia inducida por la emoción. En el caso 2, se debe a una urgencia inducida por la emoción. En el caso 3, podría deberse a cualquiera de las dos o bien a la interacción de ambas.

t_1	t_2	t_3	t_4
A	3		
B		5	
Caso 1: *impaciencia*			
A		3	
	B	4	
Caso 2: *urgencia*			
A		3	
	B		6
Caso 3: *impaciencia o urgencia*			

Cuadro 8.2

La impaciencia, un problema muy discutido, fue un tema central del Capítulo 6. Si bien la urgencia se examina menos a menudo, creo que puede ser muy importante. En particular, la urgencia de las emociones representa uno de los mecanismos mediante los cuales éstas pueden afectar la formación de creencias. Como veremos en el Capítulo 11, la formación de creencias racionales exige una recolección óptima de información. En vez de guiarse por las pruebas ya a su alcance, un agente racional recogerá pruebas adicionales antes de actuar, si la decisión que debe tomar es lo bastante importante, y el precio de esperar, suficientemente pequeño. Con frecuencia, las emociones urgentes se desatan en situaciones en las cuales el precio de esperar *es* alto, esto es, frente a un agudo peligro físico. En tales casos, es esencial actuar rápidamente sin tomarse el tiempo de averiguar

más. Pero cuando la espera puede mejorar una decisión importante, un deseo de acción inmediata inducido por la emoción puede ser nocivo. Como dijo Séneca: «La razón presta oídos a ambas partes y luego procura posponer la acción, aun la propia, a fin de ganar tiempo para escrutar la verdad; pero la ira es precipitada». El proverbio «antes de que te cases, mira lo que haces» [*«marry in haste, repent at leisure»*], sugiere tanto el ímpetu de la emoción como las consecuencias desafortunadas de la incapacidad de resistirse a ella.[73]

Antes mencioné que no todas las emociones tienen una vida media breve. De ordinario, no obstante, decaen bastante rápidamente con el tiempo. En algunos casos, esto se debe al solo hecho de que la situación que las ha provocado deja de existir. Cuando logro ponerme a buen resguardo del oso que me amenazaba, el miedo ya no se justifica. Más a menudo, sin embargo, la emoción declina a medida que el recuerdo se desvanece, por el mero paso del tiempo. La ira, la vergüenza y la culpa rara vez persisten con la intensidad que tuvieron al desatarse la emoción. Luego del 11 de septiembre de 2001, por ejemplo, el número de varones jóvenes estadounidenses que expresaron interés en servir en el ejército aumentó un 50%, pero en los hechos el reclutamiento no mostró un crecimiento pronunciado. Estos datos son coherentes con la hipótesis de que la oleada inicial de interés se debía a la emoción, luego mitigada durante los varios meses requeridos por el trámite de reclutamiento. Entre las mujeres jóvenes, el interés por alistarse casi no aumentó, un hecho que no tiene una explicación evidente.

Dicho esto, las personas son con frecuencia incapaces de prever el declive de sus emociones. Cuando están embargadas por una emoción fuerte, quizá crean, erróneamente, que ésta durará para siempre, y pier-dan incluso todo sentido del futuro. Si los individuos suicidas a quienes me he referido hubieran sabido que su vergüenza (y el desprecio de los observadores) menguaría, quizá no se habrían matado. Si las parejas jóvenes supiesen que su amor mutuo puede no ser eterno, estarían menos dispuestas a concertar compromisos vinculantes y, en particular, a hacer un «matrimonio pactado» del cual es más difícil salir.

Permítaseme concluir este punto señalando una interacción entre dos fenómenos inducidos por emociones: la inversión de preferencias y la formación de creencias nebulosas. Según una visión optimista, estos dos fenómenos podrían anularse uno a otro: debido a la inversión de

preferencias, queremos actuar en contra de nuestro juicio calmo y reflexivo, pero debido a la creencia nebulosa somos incapaces de llevar a cabo nuestra intención. Con mayor frecuencia, creo, ambos se reforzarán mutuamente. La venganza es un ejemplo. El riesgo es mínimo si no me vengo de una afrenta; mayor si me vengo pero espero mi oportunidad, y máximo si tomo venganza de inmediato sin miramiento alguno por los riesgos. Montaigne hizo una observación similar: «Quiere la filosofía que, al castigar las ofensas recibidas, evitemos la cólera: no con el fin de que la venganza sea menor, sino, por el contrario, con el fin de que la asestemos mejor y con más fuerza; para lo cual parécele que estorba esa impetuosidad». Eso es, no obstante, ignorar la paradoja de que, si no sentimos emoción, quizá no queramos venganza, y si la sentimos, tal vez no seamos capaces de llevar ésta a cabo con eficacia.

La emoción y la creencia

La emoción puede afectar la formación de creencias tanto directa como indirectamente. El efecto directo produce creencias sesgadas, y el efecto indirecto, creencias de baja calidad. Una forma de parcialidad tiene su ilustración en la teoría de la *cristalización* de Stendhal. El origen del término es el siguiente:

> En las minas de sal de Hallein, cerca de Salzburgo, los mineros echan, en las profundidades abandonadas de la mina, una rama de árbol despojada de sus hojas por el invierno; transcurridos dos o tres meses, por efecto de las aguas cargadas de partículas salinas que humedecen la rama, y luego la dejan en seco al retirarse, aparece toda cubierta de brillantes cristales. Las ramillas más chicas, no más gruesas que la pata de un mosquito, están recamadas de infinidad de cristalitos titilantes y deslumbradores.

La analogía con el amor es clara:

> En el momento en que comienza a interesarse por una mujer, ya no la ve *tal como realmente* es, sino tal como le conviene que sea. Compara las ilusiones favorables que produce ese comienzo de interés con esos preciosos diamantes que cubren la rama deshojada por el invierno, y que sólo los percibe, nótelo bien, la vista del hombre que comienza a enamorarse.

Según un proverbio francés que he citado antes y que volveré a citar, es fácil *creer lo que se teme*. También ésta es una forma de sesgo. Además del hecho de que tendemos naturalmente (aun en estados no emocionales) a dar excesiva importancia a los riesgos de baja probabilidad (Capítulo 12), los

sentimientos de miedo visceral pueden llevarnos asimismo a creer que los peligros son más grandes de lo que en realidad son. Cuando damos un paseo nocturno por el bosque, un sonido o un movimiento pueden desencadenar el miedo, y esto nos lleva a continuación a interpretar como alarmantes otros sonidos o movimientos que antes habíamos ignorado. El miedo «se alimenta de sí mismo». Quizá también sea oportuno concebir esta situación como el «efecto Otelo». Tendré más que decir al respecto en el Capítulo 23.

La urgencia de la emoción actúa más sobre la recolección de información con anterioridad a la formación de la creencia que sobre la creencia misma. El resultado es una creencia de baja calidad, basada en una cantidad de información que dista de ser óptima, pero no una creencia sesgada a favor o en contra de ninguna conclusión específica que al agente le gustaría cierta. En la práctica, sin embargo, los dos mecanismos tienden a presentarse juntos y reforzarse uno a otro. El agente se forma en principio una inclinación inducida por la emoción, y la urgencia de esta última le impide luego recoger la información capaz de corregirla. Como hemos visto en el capítulo anterior, el pensamiento desiderativo está hasta cierto punto sometido a restricciones de la realidad. También el pensamiento contradesiderativo las sufre. Por ende, si el agente hubiera recogido más información, le habría resultado difícil persistir en la creencia sesgada.

Debo reiterar, con todo, que tal vez no sea racional recoger grandes cantidades de información si el coste de oportunidad de hacerlo es elevado. De consagrar demasiado tiempo a decidir si la forma en el sendero es un palo o una serpiente, podemos llegar a terminar muertos. También debería agregarse que en los estados no emocionales, hay una tendencia a *hacer caso omiso* de los costes de oportunidad y a prestar más atención a los gastos en efectivo (Capítulo 12). Enfrentado a un posible riesgo, un agente podría disuadirse de tomar grandes precauciones debido a su coste directo, al mismo tiempo que no atribuye la importancia suficiente al hecho de que la inacción también podría ser costosa. En algunas circunstancias, por lo tanto, la urgencia de la emoción podría plantear un correctivo útil a esta tendencia irracional. Al mismo tiempo, insistamos, la urgencia y las creencias de baja calidad pueden crear problemas en vez de resolverlos.

Las emociones y la trasmutación

La jerarquía normativa de las motivaciones (Capítulo 4) puede hacer que la gente se sienta avergonzada de sus emociones. La envidia, por ejemplo, es una emoción que pocos admiten tener. Cuando Yago dice: «Si Casio sobrevive, a todas horas afeará mi vida su limpio proceder», muestra una franqueza poco común, quizá poco verosímil. La reacción más corriente a la percepción de la propia envidia es un encogimiento mental de hombros. Uno registra un sentimiento doloroso de inferioridad, tiene un fugaz deseo destructivo y luego sigue adelante. A veces, sin embargo, la emoción puede tener tanta fuerza que es imposible ignorarla. Al mismo tiempo, no es posible admitirla. La solución al conflicto consiste en que la envidia se trasmute en indignación virtuosa, por medio de una reescritura oportuna del guión. Puedo contarme una historia en la que el otro ha obtenido el objeto envidiado por medios ilegítimos e inmorales, y quizás a mis expensas, y de ese modo trasmuto la envidia en indignación aristotélica o ira.

En el Capítulo 7 examiné la formación de creencias motivadas. Como muestra la Figura 8.2, este fenómeno puede estar inmerso en el proceso de *motivación motivada* que consideré en el Capítulo 4. Para que el agente pueda adoptar una motivación de la que no se avergüence, tal vez sea necesaria una reescritura cognitiva.

Esta metamorfosis de la envidia en indignación virtuosa también puede darse de una manera más directa. La envidia no sólo es vergonzosa: es dolorosa. La creencia de que otra persona ha tenido éxito donde yo también podría haberlo tenido si me hubiera esforzado más, puede ser intensamente displacentera. Para mitigar el dolor, puedo hacer mía una historia similar acerca de las deshonrosas causas del éxito de la otra persona. La culpa también puede ser difícil de tolerar, sobre todo si nos domina la altivez. Los individuos altivos no tropezarán con dificultades para encontrar una historia que culpe a su víctima en vez de a sí mismos: «Odian también a quienes injurian» (Séneca). Este patrón parece común, por ejemplo, en los príncipes del Renacimiento, los tiranos de la Anti-güedad y otros individuos arrogantes. La preocupación por la imagen de sí mismo puede igualmente llevar a evitar la pesadumbre. En algunos individuos, el hecho de admitirse que han cometido un error es tan doloroso que preferirían dedicarse a una actividad improductiva antes que dejarlo atrás, como dictaría la racionalidad. Ésta es la «falacia del coste hundido», a veces también

llamada «falacia de Vietnam» o «falacia del Concorde». Tendré más que decir sobre ella en el Capítulo 12.

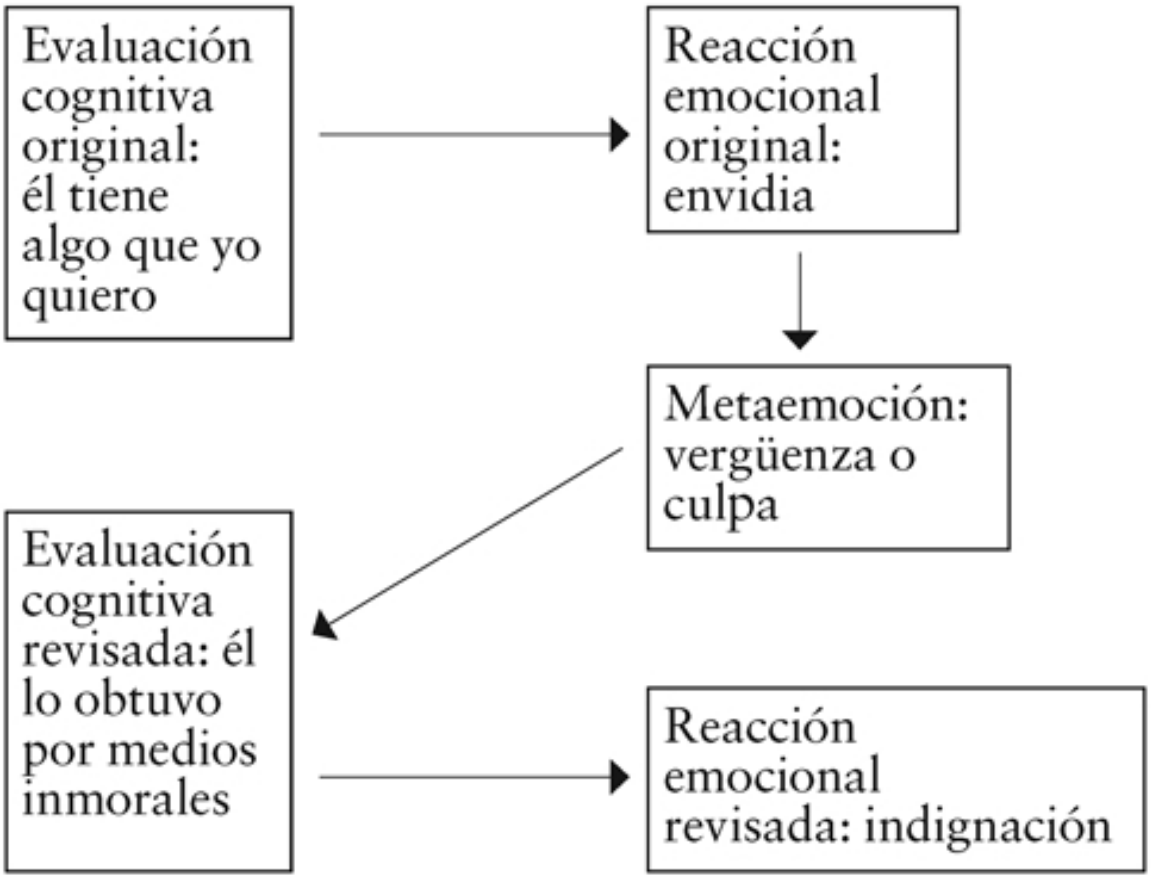


FIGURA 8.2

La cultura y las emociones

¿Son universales todas las emociones? Si no lo son, ¿hay algunas emociones universales? Respondo con un firme sí a la segunda pregunta, y con un sí tentativo a la primera.

Parece claro que algunas emociones son universales. Hay media docena de ellas, felicidad, sorpresa, miedo, tristeza, repugnancia e ira, que tienen expresiones faciales reconocidas por los miembros de todas las culturas. Si uno cree, como yo, que en todas las sociedades hay normas sociales, las emociones que les sirven de base, el desprecio y la vergüenza, también deben ser universales. Cabría imaginar una sociedad donde la gente sintiera

ira al ser ofendida, pero no indignación (cartesiana) al observar una ofensa contra un tercero. Me cuesta creer que una sociedad así pudiera existir, pero tal vez me equivoque. Si el amor es universal (véase el examen ulterior), ¿no lo serán también los celos?

Se dice que los japoneses tienen una emoción, *amae* (cuya traducción aproximada es «indefensión y deseo de ser amado»), que no existe en otras sociedades. Se ha sostenido, asimismo, que la antigua Grecia era una «cultura de la vergüenza» diferente de las «culturas de la culpa» modernas, que el amor romántico es una invención moderna y que el sentimiento de aburrimiento (si es una emoción) es de origen reciente. No se puede excluir, sin embargo, que las emociones presuntamente ausentes hayan existido sin que los integrantes de la sociedad en cuestión las conceptualizaran. Una emoción puede ser reconocida como tal por un observador externo, pero no admitida por los miembros de esa sociedad. En Tahití, un hombre cuya amiga lo haya dejado mostrará los síntomas de conducta de la tristeza, pero sólo dirá que está «cansado». En Occidente, el *concepto* de amor romántico es relativamente reciente, pues se remonta a la época de los trovadores. Con anterioridad, sólo había «alegre sensualidad o locura». Es posible, no obstante, y probable, en mi opinión, que la *experiencia* del amor romántico se produjera aun cuando la sociedad careciera del concepto para nombrarla. Los individuos pueden estar enamorados sin advertirlo, y al mismo tiempo su emoción quizá sea evidente para los observadores, pertenezcan a su propia sociedad o a otra. Los antiguos griegos exhibían un haz de reacciones relacionadas con la culpa, ira, perdón y reparaciones, que señalan la presencia de la emoción, aunque no tuvieran una palabra para designarla. La manera de pensar las emociones puede ser específica de la cultura, sin que las emociones mismas lo sean.

Debe agregarse, con todo, que cuando una emoción determinada no es explícitamente conceptualizada, es posible que también tenga menos manifestaciones de conducta. La Rochefoucauld escribió que «algunas personas nunca se habrían enamorado si nunca hubiesen escuchado hablar de amor». Del mismo modo, la culpa puede ser más común en las sociedades donde se dice a la gente, desde su más tierna infancia, que debe sentirla en tal o cual ocasión.

* * *

Nota bibliográfica

El mejor libro sobre la emoción es Nico H. Frijda, *The Emotions*, Cambridge y Nueva York, Cambridge University Press, 1986. Lo utilicé de manera sustancial en mi *Alchemies of the Mind* (cuyo subtítulo es *Rationality and the Emotions*), Cambridge y Nueva York, Cambridge University Press, 1999 [*Alquimias de la mente: la racionalidad y las emociones*, Barcelona, Paidós, 2002], donde el lector podrá encontrar más referencias a los conceptos examinados aquí. La idea de urgencia, diferenciada de la impaciencia, es una adición al marco de ese libro. Hay sostenidas discusiones del papel de las emociones en la explicación del comportamiento en Roger Petersen, *Understanding Ethnic Violence: Fear, Hatred, and Resentment in Twentieth-Century Eastern Europe*, Cambridge y Nueva York, Cambridge University Press, 2002, y en el Capítulo 8 de mi *Closing the Books: Transitional Justice in Historical Perspective*, Cambridge y Nueva York, Cambridge University Press, 2004 [*Rendición de cuentas: la justicia transicional en perspectiva histórica*, Buenos Aires, Katz, 2006].

Notas al pie

[62] Con finalidades diagnósticas, podría estipularse que un estado mental es una emoción si tiene (digamos) ocho o más de doce rasgos definitorios. Sin embargo, esta estipulación es insatisfactoria cuando el propósito consiste en explicar.

[63] Es posible, con todo, que pongan parcialmente en juego el mismo circuito neuronal.

[64] Las emociones provocadas por las evaluaciones negativas de uno mismo siempre tienen una valencia negativa. Las causadas por las evaluaciones negativas de otros son más ambiguas en este aspecto.

[65] El primero en identificar la emoción fue Descartes, que añadió la importante salve-dad de que, cuando el agente *ama* al tercero, la reacción es de ira y no de indignación.

[66] Incluyo «no inmerecida» bajo el encabezado de «merecida». Así, cuando alguien gana el gordo en la lotería, diré que es merecido, un uso un tanto contrario al habitual.

[67] Aunque el término utilizado por Aristóteles para designar esta emoción suele traducirse como «indignación», deberían resultar claras sus diferencias con respecto a la indignación cartesiana.

[68] Las emociones también pueden ser generadas, quizá de manera menos habitual, por la incertidumbre acerca del pasado. Tras la caída de la República Democrática Alemana, muchos ciudadanos solicitaron acceso a sus archivos de seguridad para averiguar si algunos de sus amigos y parientes habían suministrado informaciones sobre ellos. Lo más conmovedor es que algunos querían saber si sus fracasos personales (no haber obtenido un ascenso o haber sido abandonados por un amante) se debían a sus propias flaquezas o a las instrucciones dadas a otras personas por los servicios de seguridad del Estado para aislarlos.

[69] Los «tiros errados por poco» suscitan emociones más fuertes; así, los medallistas de plata de las competencias olímpicas se manifiestan menos felices que los medallistas de bronce.

[70] Supongamos que la probabilidad previa de una victoria es p y que la satisfacción obtenida con ella es proporcional a $1/p$ (porque la sorpresa es más grande cuando p es baja). En este modelo especial de la sorpresa, la satisfacción *esperada* de la victoria es independiente de la probabilidad de ésta. En general, el impacto de la sorpresa va a ser más complejo.

[71] Sospecho que las transgresiones de las normas cuasimorales provocan las mismas emociones que las transgresiones de las normas morales, pero mi intuición no es sólida.

[72] Así, en respuesta al asesinato de trescientos sesenta y cinco musulmanes libaneses, una libanesa dijo que «en este momento quiero que [la milicia musulmana] […] vaya a las oficinas y mate a los primeros setecientos treinta cristianos indefensos a los que pueda meter mano». En *La guerra civil en Francia*, Marx cita una carta publicada en un diario, en la que un sacerdote tomado como rehén por los comuneros dice que éstos «han tomado la decisión de ejecutar a dos de los numerosos rehenes que tienen en su poder luego de cada nueva ejecución» (perpetrada por las fuerzas gubernamentales). Séneca, para terminar, señaló que «un mal no excedido no queda vengado».

[73] El cercenamiento de las libertades civiles impuesto por varios Gobiernos occidentales a raíz de los ataques del 11 de septiembre de 2001 es un buen caso testigo. ¿Ilustró esa actitud la necesidad de actuar con rapidez en presencia de un peligro inminente, o fue una reacción de pánico que, al hacer que esos Gobiernos aparecieran aún más aborrecibles a los ojos de sus enemigos, aumentó la probabilidad de nuevos ataques en vez de disminuirla? Véanse también las observaciones sobre la «psicología de la tiranía» en capítulos anteriores.

TERCERA PARTE

Acción

Si bien haré amplio uso de los términos «acción», «comportamiento», «decisión» y «elección» como sinónimos, a veces es útil distinguir entre ellos. La categoría más general es la del *comportamiento*, entendido como cualquier movimiento corporal cuyo origen es interno al agente, no externo (como lo es cuando lo arrastra una avalancha). La *acción* es un comportamiento intencional, causado por los deseos y las creencias del agente. Así, los comportamientos reflejos no son acciones; tener una erección no es una acción (pero puede ser inducido por una, como el haber tomado Viagra); dormirse no es una acción (pero puede inducirlo la toma de un somnífero). Una acción puede ser precedida, o no, por una *decisión* consciente. Cuando voy en automóvil al trabajo por el camino habitual, no decido conscientemente girar a la derecha aquí y a la izquierda allí, aunque cada una de esas acciones sea intencional o esté orientada hacia un objetivo. Sin embargo, la primera o las primeras veces que utilicé el auto para ir al trabajo, las acciones fueron precedidas por decisiones explícitas. De hecho, las precedió una *elección* explícita entre caminos alternativos. Si bien todas las elecciones son decisiones, lo inverso no es cierto. Cuando decido tomar el libro que he estado leyendo, no es necesario que tenga una alternativa explícita en mente. Veo el libro sobre la mesa; su visión me recuerda que he disfrutado de su lectura, y decido entonces tomarlo. No interviene ninguna elección.

La característica más importante de este paisaje conceptual es tal vez el hecho de que *no todas las decisiones conducen a acciones*. Uno puede decidir no hacer algo, por ejemplo, no salvar a una persona que se está ahogando, si la intervención implica algún riesgo para uno mismo. Si la persona se ahoga y no hay terceros involucrados, no tengo una responsabilidad causal por el resultado. Quizá tenga una responsabilidad moral y, en algunos países, una responsabilidad legal, pero eso es harina de otro costal.[1] Supongamos, no obstante, que hay un tercero presente o, como en el caso de Kitty Genovese, muchos terceros. Si el tercero observa que yo estoy en condiciones de ayudar a la persona que se ahoga y no lo hago, podría razonablemente inferir que la situación es menos grave de lo que habría parecido y, en consecuencia, también se abstendrá de intervenir. En ese caso, mi decisión de no hacer nada habría llevado a otra persona a decidir en el mismo sentido. De tal modo, las decisiones pueden tener eficacia causal aun cuando no generen una acción.

La tercera parte se organiza en casi toda su extensión alrededor de la teoría de la elección racional. Como he explicado en la introducción, soy ahora más escéptico que antes con respecto a las explicaciones de la acción (¡o de la inacción!) basadas en esa teoría. Sin embargo, aunque gran parte del comportamiento es irracional, de una manera u otra, la racionalidad sigue siendo primordial en cierto sentido. Los seres humanos *quieren* ser racionales. No nos enorgullecen nuestras caídas en la irracionalidad. Antes bien, tratamos de evitarlas o corregirlas, a menos que nuestra altivez nos impida reconocerlas.

Nota

[1] En los Estados Unidos no existe la obligación de ser un buen samaritano, excepto en condiciones muy circunscritas. En la Europa continental, la «falta de asistencia a una persona en peligro» puede ser severamente castigada si el riesgo corrido por el samaritano al ayudarla es pequeño en comparación con el de quien está necesitado de ayuda. Algunos jurisconsultos estadounidenses sostienen que la ley de los Estados Unidos es más eficiente, dado que una obligación general de prestar asistencia crearía un incentivo para que potenciales rescatistas evitaran los lugares donde existe la probabilidad de que se necesite un rescate, debido a la amenaza de ser responsables. Esto puede ser así o, más probablemente, no serlo; lo que sí parece indudable es que el sistema estadounidense no surgió por razones de eficiencia, y, tras establecerse por otras razones, tampoco se mantuvo en vigencia por sus propiedades en esa materia.

Capítulo 9
Deseos y oportunidades

Hacer todo lo posible

Para caracterizar el comportamiento, a veces decimos: «Hizo todo lo posible». Si desarmamos esta frase, veremos que contiene dos elementos: deseos y oportunidades. Los deseos definen lo que el agente considera «lo posible». Las oportunidades son las opciones o medios entre los que el agente «puede» elegir. Esta caracterización quizá sirva también como *explicación* rudimentaria del comportamiento sobre la base de la elección racional. Si preguntamos: «¿por qué lo hizo?», la respuesta «era lo mejor que podía hacer» tal vez sea plenamente suficiente. En muchos casos, se necesita algo más para proporcionar una explicación satisfactoria en función de una elección racional. En particular, acaso necesitemos recurrir a las *creencias* del agente y no sólo a los deseos y oportunidades. Nos ocuparemos de estas complicaciones en el Capítulo 11. Aquí, examino hasta dónde puede llevarnos el solo marco de los deseos y las oportunidades.[2] También sugeriré que el marco no es a veces tan simple como parece, debido a que los deseos y las oportunidades no siempre son independientes entre sí, como en ocasiones se supone.

Hay otra manera equivalente de contemplar el asunto. En la tarea de comprender el comportamiento, podemos comenzar con todas las acciones posibles en abstracto, que el individuo podría realizar. La acción que observamos verdaderamente puede considerarse como el resultado de dos operaciones sucesivas de filtrado. El primer filtro está constituido por todas

las *restricciones*, físicas, económicas, legales y otras, que enfrenta el agente. Las acciones compatibles con todas ellas constituyen el conjunto de oportunidades.[3] El segundo filtro es un mecanismo que determina cuál de las acciones del conjunto de oportunidades se ha de llevar realmente a cabo. Aquí, supongo que el agente escoge la acción que promete las mejores consecuencias, conforme lo evalúan sus deseos (o preferencias). En capítulos siguientes consideraremos otros mecanismos que actúan como segundo filtro.

El enfoque de los filtros sugiere la siguiente pregunta: ¿qué sucede si las restricciones son tan fuertes que no hay ningún elemento sobre el que el segundo filtro pueda actuar? ¿Puede ser que las restricciones determinen de manera singular una acción, y una sola, compatible con todas ellas? Tanto los ricos como los pobres tienen la oportunidad de dormir bajo los puentes de París, pero quizá los pobres carezcan de otras opciones.[4] En el caso de un consumidor pobre, las restricciones económicas y calóricas podrían determinar en conjunto un haz único de bienes.[5] Puede interpretarse que quienes defienden la idea del *estructuralismo* en las ciencias sociales digan que, de ordinario, las restricciones son tan fuertes que dejan poco o ningún campo a la elección.[6] Sin embargo, sigue siendo un misterio *por qué* debería ser así. Uno no puede sostener, por ejemplo, que los ricos y poderosos se aseguren de que los pobres y oprimidos no tengan otra opción que trabajar para ellos, puesto que este enunciado presupone que los primeros sí tienen, al menos, una opción.

En algunos casos, el enfoque de los deseos y las oportunidades es incompleto. Quizá yo tenga la oportunidad de realizar la acción más adecuada para cumplir mi deseo, como escoger la respuesta correcta en una prueba o dar en el blanco en una competencia de tiro con rifle, pero no la *aptitud* de identificar qué acción se ha de tomar. En ciertos casos podríamos situar esa incapacidad en una etapa anterior, en la que el agente carecería o bien de la oportunidad o bien del deseo de adquirir la aptitud. En otros, la incapacidad se deduce de duras restricciones psicofisiológicas. Puedo tener tanto el deseo como la oportunidad de escoger la acción que maximice mi bienestar a largo plazo, pero no la aptitud de determinar, en el acto, cuál es esa acción. Cuando los economistas y los politólogos de inclinaciones matemáticas tratan de establecer cuál es el comportamiento óptimo en una situación dada, a menudo necesitan muchas páginas de complicadas fórmulas para explicarlo en detalle. Algunos individuos tal vez nunca

puedan, por mucha preparación que reciban, ser capaces de realizar esos cálculos o no puedan realizarlos sin esforzarse más de lo que la tarea merece o más tiempo del que tienen a su disposición.

Explicaciones basadas en la oportunidad

Aun cuando el comportamiento es el resultado conjunto de deseos y oportunidades, su *varianza* a lo largo del tiempo puede explicarse en gran medida sobre la base de las segundas. El consumo de alcohol está determinado, en general, por la fuerza del deseo de beber, comparándolo con otros deseos, y por lo que el bebedor puede solventar. Cuando los precios del alcohol sufren un alza pronunciada, por ejemplo en épocas de guerra, el consumo cae abruptamente.

Se puede plantear una explicación con las curvas de indiferencia (Figura 9.1). Supongamos que el consumidor tiene que distribuir sus ingresos entre el alcohol y algún lote de bienes de consumo corrientes. Los precios relativos y sus ingresos son en un comienzo tales que la persona enfrenta el conjunto de oportunidades situado dentro del triángulo OAA'. Suponiendo que gasta todo su ingreso, podemos limitarnos a la línea del presupuesto AA'.[7] La fuerza de su deseo de alcohol en comparación con el lote de bienes de consumo se muestra en la forma de las *curvas de indiferencia* I, I' e I''. El término refleja la idea de que al consumidor le son indistintas todas las combinaciones de alcohol y otros bienes que se dan en cualquier curva dada, a la vez que prefiere cualquier combinación en una curva más alta a cualquier combinación en una curva más baja.[8] Para elegir la mejor entre las opciones a su alcance, el consumidor debe escoger el punto de la línea del presupuesto que es tangente a una curva de indiferencia, pues ésta será la más elevada entre las curvas que incluyen una combinación que él puede permitirse. En la Figura 9.1, esto produce el consumo de alcohol OX.

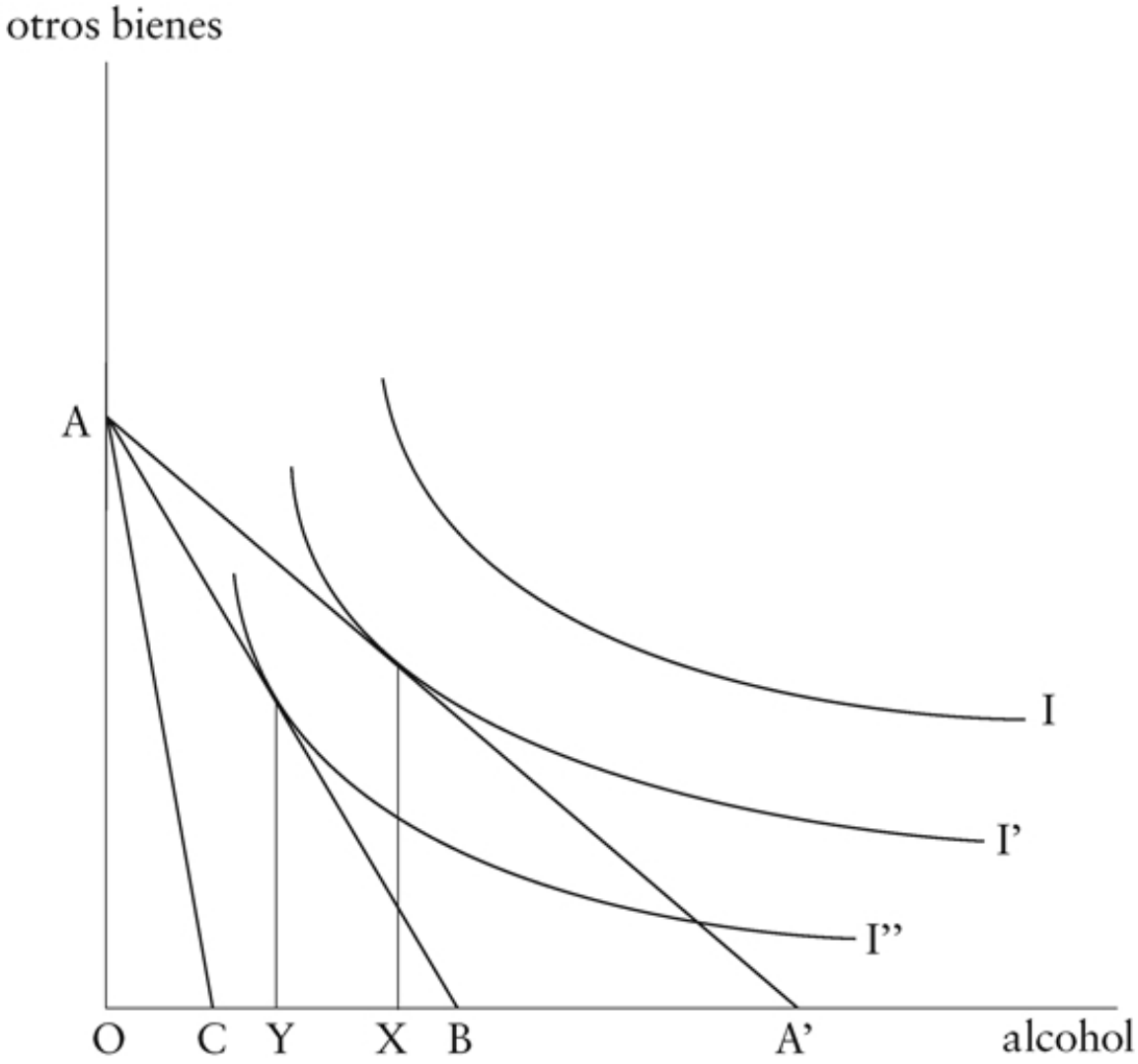


FIGURA 9.1

Supongamos ahora que el precio del alcohol sube a tal punto que el consumidor se ve frente a la línea de presupuesto AB. Como el punto de tangencia se ha desplazado hacia la izquierda, el consumidor consumirá ahora OY. Podríamos llevar hasta su término el mismo razonamiento si un nuevo aumento de precios trasladara la línea de presupuesto a AC. No obstante, aunque no sepamos nada sobre la forma de las curvas de indiferencia, podemos predecir que en esta situación el consumidor no consumirá más que OC, como sucedería si gastara todos sus ingresos en alcohol. El conjunto de oportunidades puede explicar por sí mismo gran parte de la varianza a lo largo del tiempo. El segundo filtro, de hecho,

podría ser cualquier cosa, un comportamiento optimizador, una irresistible avidez de alcohol, la costumbre o lo que fuera, y aun así el consumidor estaría seriamente limitado por el primero.

Escojo este ejemplo en particular para analizar la cuestión de los deseos presuntamente «irresistibles», como el de los drogadictos, los grandes fumadores o los alcohólicos por la sustancia a la que son adictos. ¿Las drogas se asemejan más a la insulina, que un diabético comprará a cualquier precio, o al azúcar, que los consumidores comprarán regularmente en menor cantidad cuando los precios aumenten? Como prueba de que se asemejan más al azúcar, se menciona el hecho de que el consumo de drogas baja cuando los precios suben. Sin embargo, como hemos visto, este fenómeno podría deberse simplemente a la incapacidad del adicto de consumir por encima de su presupuesto. (El diabético también podría ser incapaz de comprar la insulina que necesita si el precio se incrementa.) Así, parece que la caída del consumo de alcohol en épocas de guerra se debe con frecuencia a su inasequibilidad, lo cual deja sin resolver la cuestión de los deseos irresistibles en comparación con los deseos resistibles. No obstante, sí sabemos con otros fundamentos que el consumo de alcohol *es* sensible a los precios. Aun cuando los consumidores puedan darse el lujo de mantener su nivel previo de consumo a precios más altos, no lo hacen.

En ocasiones se propone otro argumento para mostrar que las oportunidades tienen mayor capacidad explicativa que los deseos. Algunos economistas afirman que todos los individuos tienen en esencia los mismos deseos y preferencias: sólo difieren las oportunidades. Aunque parezca que el gusto de la gente en materia de música clásica difiere, las diferencias de consumo observadas se deben simplemente (así se sostiene) al hecho de que algunos individuos tienen más «capital de consumo musical» y, por ende, se complacen más que otros con ese tipo de música. Un oído entrenado tiene más que ver, a buen seguro, con una aptitud que con una oportunidad, pero se trata de una aptitud que depende de la oportunidad de adquirirla. Y se supone que el deseo de adquirirla es el mismo en todos. Esta última afirmación, sin embargo, revela una debilidad en el argumento. En nuestros días, el acceso a la música clásica en la radio es esencialmente gratuito, aun cuando se tomen en cuenta los costes de oportunidad (podemos escuchar mientras hacemos otras cosas). Como no todo el mundo desea tener un oído entrenado, la explicación debe encontrarse en la inclinación subjetiva a

adquirirlo, que es una función (entre otras cosas) del modo como la música clásica llega a distintos oídos *no entrenados*.

Otros aspectos en que las oportunidades son más importantes que los deseos

Las oportunidades son más básicas en un aspecto fundamental: son más fáciles de observar, no sólo por los especialistas en ciencias sociales, sino también por otros integrantes de la sociedad. Una de las máximas básicas de la estrategia militar sostiene que se debe planificar en función de las oportunidades (verificables) del adversario, no en función de sus intenciones (inverificables).[9] Si tenemos razones para creer que el adversario *podría* tener intenciones hostiles, la máxima puede llevarnos a suponer lo peor: si puede, nos perjudicará.[10] La situación se complica por el hecho de que nuestra creencia en las intenciones hostiles del adversario puede fundarse en la percepción de que *él* cree que tenemos los medios y quizá la intención de agredirlo. En esta ciénaga de subjetividad, las oportunidades objetivas tal vez proporcionen la única base sólida para la planificación.

Otra razón por la cual las oportunidades pueden parecer más fundamentales que los deseos tiene que ver con la posibilidad de influir sobre el comportamiento. De ordinario es más fácil modificar las circunstancias y oportunidades de la gente que modificar sus opiniones.[11] Éste es un argumento en términos de costes y beneficios sobre la eficacia monetaria de las políticas alternativas, no sobre la capacidad explicativa relativa. Aunque el gobierno tenga una buena teoría, con cabida tanto para la explicación como para la predicción, quizás ésta no le permita ejercer mucho control, vista la posibilidad de que los elementos sobre los que puede actuar no sean los de importancia causal. Supongamos que las causas de un mal desempeño económico pueden encontrarse en la existencia de sindicatos fuertes y de empresarios poco dispuestos a arriesgarse. El gobierno tal vez esté plenamente convencido de que la actitud mental de los administradores de empresas es la causa más importante, no obstante lo cual sea incapaz de hacer nada al respecto. En contraste, como han mostrado los años de Reagan y Thatcher, la acción gubernamental puede quebrar la oposición de los sindicatos.

Un ejemplo importante, digno de considerarse, es el comportamiento suicida. Para suicidarse, no basta con el deseo de matarse: también es preciso encontrar los medios de hacerlo. El alto índice de suicidios entre los médicos, por ejemplo, puede deberse en parte a su fácil acceso a drogas letales, que son en este grupo el método predilecto para quitarse la vida.[12] Aunque el Gobierno trate de limitar las intenciones suicidas, y para ello establezca líneas telefónicas de ayuda o persuada a los medios de comunicación de restringir los informes sobre suicidios, que pueden suscitar un contagio, los resultados más efectivos se obtienen cuando se dificulta el acceso a los instrumentos para concretarlo.[13] Entre las medidas tomadas se incluyen barreras que hacen más difícil saltar de los puentes o los edificios altos, el control más riguroso de ciertos medicamentos de venta bajo receta, restricciones a la venta de armas cortas, el reemplazo del letal monóxido de carbono por el gas natural en los hornos de las cocinas y la instalación de convertidores catalíticos que reducen las emisiones de ese mismo monóxido de carbono en los tubos de escape de los automotores. En el futuro, tal vez veamos la prohibición de los sitios de Internet dedicados al «suicidio asistido». Aun el simple reemplazo de los frascos por blísteres ha contribuido a la reducción del número de suicidios por envenenamiento con paracetamol. La disminución de la cantidad máxima de comprimidos en las preparaciones o recetas individuales también puede reducir la probabilidad de envenenamientos graves. A diferencia del Reino Unido, Francia ha limitado por ley el contenido de cada envase de paracetamol a ocho gramos. Se estima que ésta es una razón por la cual el daño hepático grave y las muertes por envenenamiento con este fármaco son menos comunes en ese país que en Gran Bretaña.

Está claro que un individuo resuelto encontrará habitualmente un camino. Cuando se suprime el acceso a un método corriente de atentado contra la propia vida, la caída resultante en el índice de suicidios puede ser hasta cierto punto temporaria. Sin embargo, al menos en algunos casos, el efecto parece haber sido duradero, como cabría esperar. Si el impulso de matarse es fugaz y no tiene bases sólidas, podría desaparecer para el momento en que uno se las ha arreglado para hacerse con un medio adecuado.[14] Por tanto, la mera *postergación* (y no el bloqueo) del acceso a los medios podría ser eficaz para impedir los suicidios impulsivos. La exigencia de un plazo de espera antes de la compra de un arma corta podría reducir los índices tanto de suicidio como de homicidio.[15]

Variedades de las interacciones entre deseos y oportunidades

Para tomar un ejemplo más complejo de las interacciones entre deseos y oportunidades podemos recurrir al análisis de las facciones que hace James Madison en *The Federalist*. Madison sostenía que en el régimen que prevalecería en una democracia directa (que sería necesariamente reducida) o en una pequeña república representativa, las *facciones* tendrían tanto el motivo como los medios de cometer agravios. Por un lado, «la mayoría del total sentirá en casi todos los casos una pasión o un interés comunes». Por otro, la limitada cantidad de ciudadanos brinda a la mayoría la oportunidad de actuar de manera opresiva contra la minoría, por el hecho de que puede reunirse con mayor facilidad. En contraste, en una república grande es «menos probable que una mayoría del total tenga un motivo común para usurpar los derechos de otros ciudadanos; o, si ese motivo común existe, será más difícil para todos los que lo sienten descubrir su fuerza y actuar al unísono». En esta trasposición del plano individual al plano colectivo, el argumento de los deseos y las oportunidades adopta una forma un tanto diferente. Si bien el plan institucional, creía Madison, es ineficaz para modificar los motivos individuales, puede cambiar la probabilidad de que una mayoría de los individuos comparta el mismo motivo. Aunque dicho plan institucional no puede, estrictamente hablando, afectar la oportunidad de los miembros de una mayoría facciosa (en caso de que, pese a todo, la haya) de actuar de consuno, sí puede reducir su *capacidad* de hacerlo, pues disminuye la probabilidad de que cada uno sepa de la existencia de los otros. Con anterioridad a la era de las encuestas de opinión, debe haber habido, de hecho, innumerables ocasiones en las que una mayoría silenciosa omitió reconocerse como tal.

El argumento de Madison es, por así decirlo, de doble efecto: una república grande no sólo impedirá el surgimiento de mayorías facciosas, sino la acción concertada. En *La democracia en América*, de Tocqueville, encontramos una gran cantidad de argumentos con esta estructura de «no sólo». Consideremos, por ejemplo, su examen de la influencia de la esclavitud sobre los propietarios de esclavos. Ante todo, la esclavitud es improductiva si se la compara con el trabajo libre. «El obrero libre recibe un salario; el esclavo una educación, alimentos, cuidados y vestidos; el dinero que gasta el amo para el mantenimiento del esclavo se derrama poco a poco y en detalle; apenas se nota; el salario que se da al obrero se entrega

de una sola vez, y parece no enriquecer sino a quien lo recibe; pero, en realidad, el esclavo ha costado más que el hombre libre, y sus trabajos han sido menos productivos.» Sin embargo, «la influencia de la esclavitud se extiende todavía más lejos; penetra hasta en el alma misma del amo, […] imprime una dirección particular a sus ideas y a sus gustos». Como el trabajo se asocia a la esclavitud, los blancos del sur desprecian no «solamente el trabajo, sino todas las empresas que el trabajo hace prosperar». Carecen tanto de las oportunidades como del deseo de enriquecerse: «La esclavitud no impide […] solamente a los blancos hacer fortuna, sino que los desvía de intentarlo». Si Tocqueville tiene razón, el debate clásico sobre el estancamiento económico de las sociedades esclavistas es espurio. No hace falta preguntarse si la explicación correcta es la falta de deseos de inversión o la falta de oportunidades de inversión: ambas cosas podrían ser válidas.

Los argumentos planteados por Madison y Tocqueville tienen una estructura común: una misma tercera variable da forma tanto a los deseos como a las oportunidades, que en conjunto dan forma a la acción (o la impiden, según corresponda). En abstracto, hay cuatro posibilidades (los signos más y menos indican efectos causales positivos y negativos, respectivamente) (véase la Figura 9.2).

El caso (A) tiene su ilustración en el análisis de Madison de las democracias directas o las repúblicas pequeñas. El caso (B) se ejemplifica en su argumento a favor de las repúblicas grandes y en el análisis de Tocqueville de los efectos de la esclavitud sobre los propietarios de esclavos.

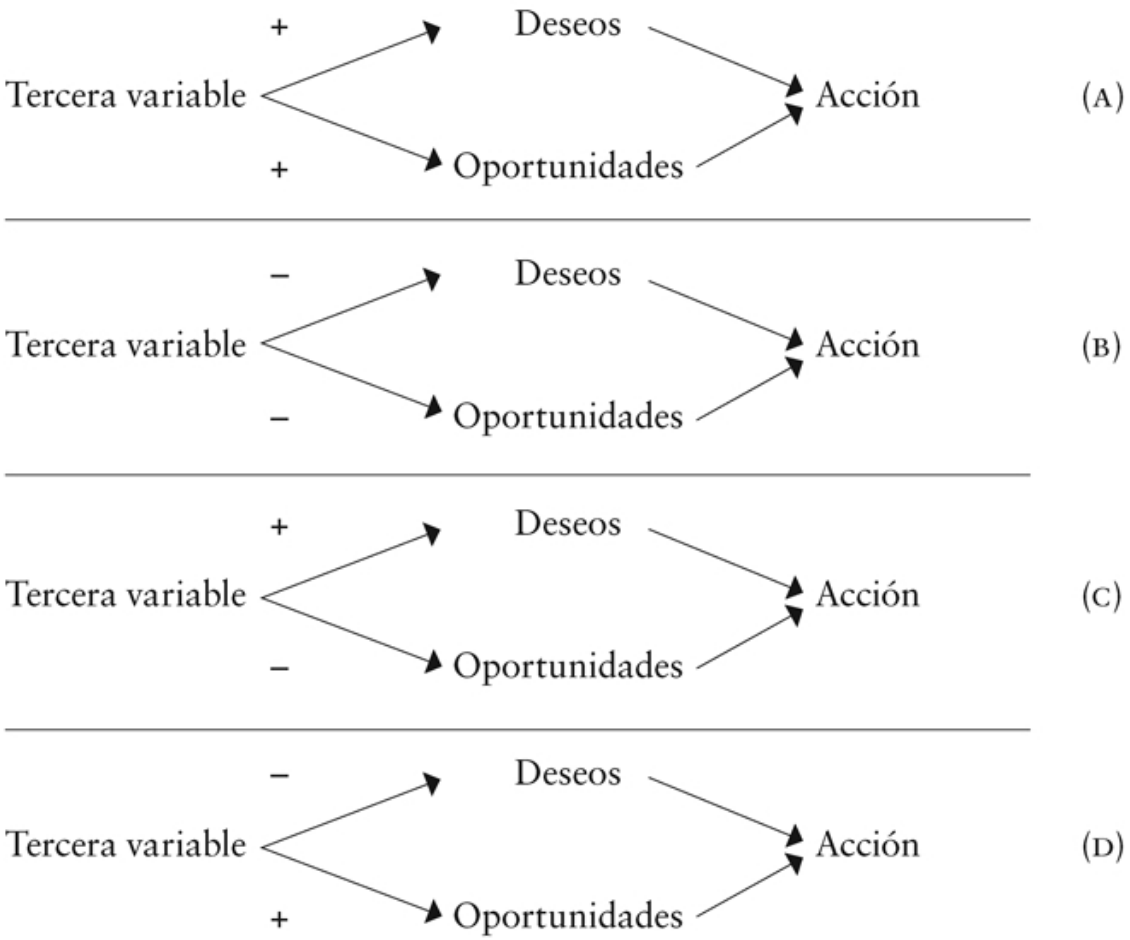


FIGURA 9.2

El caso (C) se observa en las muchas instancias en que la falta de recursos tiene el efecto dual de aumentar el incentivo para mejorar la propia situación y reducir la oportunidad de hacerlo. Aunque se dice que «la necesidad es la madre del ingenio», esto sólo es cierto en la medida en que la penuria incrementa la motivación para innovar. Pero como la innovación a menudo exige recursos (que, en consecuencia, podrían calificarse de «el padre del ingenio»), por sí misma la motivación quizá no lleve a ninguna parte. La innovación suele requerir costosas inversiones con un retorno incierto y demorado, pero eso es exactamente lo que las empresas al borde de la quiebra no pueden permitirse. Las empresas prósperas pueden darse el lujo de innovar, pero tal vez no se molesten en hacerlo. Como dijo el economista John Hicks: «La mayor de todas las ganancias monopólicas es una vida tranquila».

De manera similar, si bien el deseo de emigrar se fortalece cuando la pobreza impera en el país natal, esa misma pobreza puede bloquear el acceso a los medios de la emigración, debido a los costes del viaje. Hasta comienzos del siglo XIX, los emigrantes a los Estados Unidos podían usar su cuerpo como garantía. Sus futuros empleadores solían pagar el viaje a cambio de un período de servidumbre por contrato. En nuestros días, los contrabandistas de seres humanos pueden apoyarse en el miedo que suscita el Servicio de Inmigración y Naturalización para impedir que los inmigrantes ilegales renieguen de su promesa de reembolsar los costes del viaje con sus ingresos laborales en el país receptor. Pero cuando los irlandeses huyeron del hambre en la década de 1840, los más pobres se quedaron en su país para morir allí.

Otro ejemplo del caso (C) se encuentra en el estudio de las rebeliones campesinas: aunque los campesinos pobres tengan los mayores incentivos para rebelarse, tal vez carezcan de los recursos para hacerlo. La participación en una acción colectiva exige la aptitud de restar tiempo a las actividades productivas, que es precisamente lo que el campesino empobrecido no puede permitirse. Los campesinos medios que se las han ingeniado para ahorrar algo pueden darse el lujo de adherir a una rebelión, pero su motivación es menos aguda. Marx sostenía que la civilización surgió en las zonas templadas porque sólo en ellas el deseo de mejorar se alió a las oportunidades de hacerlo. Donde la naturaleza es demasiado pródiga no hay deseo, y donde es demasiado avara no hay oportunidades. Según muestra este ejemplo, puede haber una gama de recursos dentro de la cual tanto los deseos como las oportunidades estén lo bastante desarrollados para dar pábulo a la acción, pero a priori nada puede decirse de la amplitud o estrechez de esa gama, y ni siquiera es posible asegurar que existe.

Hemos visto un ejemplo del caso (D) en el Capítulo 2. La parte superior de la Figura 2.1 muestra el argumento de Tocqueville de que la democracia (por intermedio de la religión) inhibía el deseo de adoptar el desordenado comportamiento para cuyo ejercicio brindaban una oportunidad instituciones democráticas como la libertad de prensa y la libertad de asociación. Una observación más trivial del mismo Tocqueville se apoya en la conjunción de (C) y (D), con la juventud o la vejez como tercera variable: «En los Estados Unidos, la mayor parte de los ricos comenzaron siendo pobres. Casi todos los ociosos han sido, en su juventud, gente ocupada, de donde resulta que, cuando se podría tener el gusto del estudio, no se tiene

tiempo para dedicarse a él, y cuando se ha conquistado el tiempo para consagrarse a él, ya no se cuenta con el gusto de hacerlo».[16]

Los deseos y las oportunidades también pueden afectarse unos a otros directamente: piénsese en el primer caso, (E), de la Figura 9.3. En el Capítulo 2 me referí a algunos de los aspectos en que las oportunidades pueden afectar los deseos: la gente tal vez termine por desear más lo que puede conseguir o preferir lo que tiene a lo que no tiene. Podemos volver a citar a Tocqueville sobre la esclavitud: «¿Llamaré gracia de Dios a una maldición de su cólera, a esa disposición del alma que hace al hombre insensible a las miserias extremadas y a menudo aun le da una especie de placer depravado por la causa de sus desgracias?» Este mecanismo sugiere una razón más para concebir que las oportunidades son más básicas que las preferencias. En conjunto, las oportunidades y los deseos son las causas próximas de la acción, pero a mayor distancia sólo las primeras importan, dado que también dan forma a los segundos. El mecanismo de «formación de preferencias adaptativas» (una forma de reducción de la disonancia) garantiza que ninguna opción no perteneciente al conjunto de oportunidades sea preferida a la opción predilecta dentro de él.

Podríamos preguntarnos si este mecanismo *importa* para el comportamiento, puesto que, por definición, las opciones que no están en el conjunto de oportunidades no serán escogidas.[17] Supongamos que en un principio el agente clasifica las opciones en el orden A, B, C, D, y luego se entera de que A no está a su alcance. Mediante la formación de preferencias adaptativas, ahora las clasifica en el orden B, A, C, D. Escogerá B, como habría hecho si sus preferencias hubiesen seguido siendo las mismas. Supongamos, empero, que la nueva clasificación es C, B, A, D, lo cual induce la elección de C. Esto podría ocurrir por obra de un proceso de «sobreadaptación» a las oportunidades limitadas. Tocqueville afirmó que ésta era una característica peculiar del francés: «Va más allá del espíritu de servicio tan pronto como éste lo ha embargado». Más probablemente, estamos ante una tendencia general observada en muchas sociedades de estatus. La nueva clasificación de las preferencias también podría ser B, C, D, A. Si las mujeres hermosas rechazan mis avances, tal vez me consuele con la idea de que, en virtud de su narcisismo, son en realidad las compañeras menos deseables.

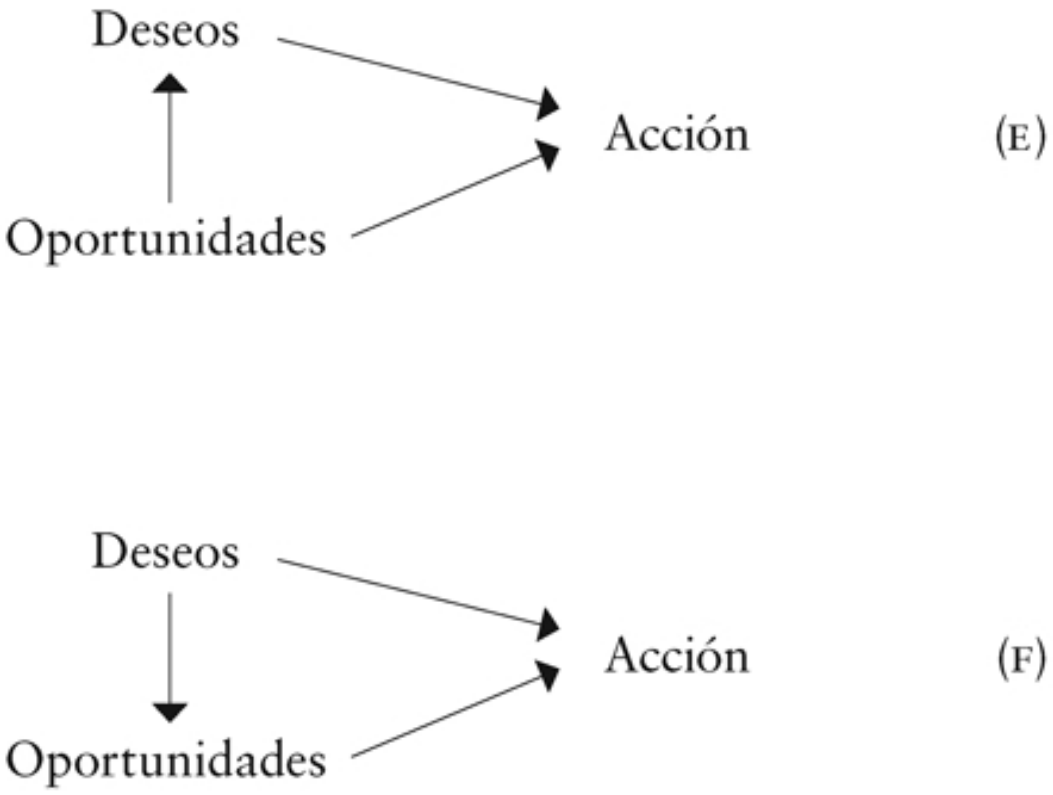


FIGURA 9.3

Examinemos por último el caso (F), en el cual el deseo del agente da forma al conjunto de oportunidades. Este resultado puede producirse sea por elección intencional, que considero en el Capítulo 13, sea por un mecanismo causal no intencional. Como he señalado en el Capítulo 4, el deseo de estados que son en esencia subproductos puede interferir en la oportunidad de provocarlos. El mecanismo podría ser intrapsíquico, como en el deseo contraproducente de dormirse, o interpersonal, como en el siguiente ejemplo. En un departamento universitario donde dicté clases, había una regla tácita según la cual cualquier miembro del cuerpo docente que mostrara notorio interés en llegar a ser presidente del departamento arruinaría con ello su oportunidad de obtener el cargo. En su trayectoria, Tocqueville ilustró ambos lados de esa proposición. En una primera etapa, sus ambiciones políticas se frustraron porque era demasiado visiblemente ambicioso (como dice el refrán, «quien entra al cónclave como papa, sale como cardenal»). Más adelante, al comentar su triunfo en las elecciones a la

Asamblea Constituyente de 1848, escribió que «nada contribuye más al éxito que el no desearlo con demasiado ardor».

* * *

NOTA BIBLIOGRÁFICA

La idea de «deseos irresistibles» es eficazmente demolida por Gary Watson, «Disordered appetites: addiction, compulsion, and dependence», en Jon Elster (comp.), *Addiction: Entries and Exits*, Nueva York, Russell Sage, 1999, págs. 3-28. La idea de que los individuos tienen las mismas preferencias y sólo difieren en las oportunidades que enfrentan se asocia sobre todo a George Stigler y Gary Becker, «De gustibus non est disputandum», *American Economic Review*, 67(2), 1977, págs. 76-90. La idea de las preferencias adaptativas y en especial de sobreadaptación a las restricciones se debe a Paul Veyne, *Le Pain et le cirque: sociologie historique d'un pluralisme politique*, París, Seuil, 1976 (hay una traducción parcial al inglés: *Bread and Circuses: Historical Sociology and Political Pluralism*, Nueva York, Penguin, 1982). La caída del índice (global) de suicidios que siguió al reemplazo del gas de carbón por el gas natural en Gran Bretaña está documentada en Norman Kreitman, «The coal gas story: United Kingdom suicide rates 1960-1971», *British Journal of Social and Preventive Medicine*, 30, 1976, págs. 86-93. Christopher Cantor y Pierre Baume, «Access to methods of suicide: what impact?», *Australian and New Zealand Journal of Psychiatry*, 32, 1998, págs. 8-14, examinan la influencia de la reducción del acceso a un método de suicidio sobre el mayor uso de otros métodos. La utilización que hace Madison de la distinción entre oportunidades y deseos se analiza en Morton White, *Philosophy*, The Federalist*, and the Constitution*, Nueva York, Oxford University Press, 1987. Examino el uso de la distinción por parte de Tocqueville en el Capítulo 4 de *Political Psychology*, Cambridge y Nueva York, Cambridge University Press, 1993 [*Psicología política*, Barcelona, Gedisa, 1995].

Notas al pie

[2] Al hacerlo, adopto la premisa implícita de que el conjunto de opciones que el agente *cree* tener a su alcance coincide con el conjunto de oportunidades «objetivas».

[3] Al hablar de restricciones legales, no me refiero al efecto de las leyes cuando consiguen que algunas acciones sean más costosas que otras (aspecto correspondiente al segundo filtro), sino a su efecto respecto de lo que es posible o imposible. Sólo puedo votar el día fijado para las elecciones, y no puedo casarme al margen de las jurisdicciones legalmente establecidas.

[4] Si el conjunto de oportunidades se describe con mayor atención al detalle, es posible que los pobres tengan la opción de elegir el puente debajo del cual quieren dormir. Ésta es una observación general: en toda descripción de las opciones, tal vez seamos capaces de especificar una situación en la cual el agente tiene una sola opción factible, pero para cualquier situación podemos encontrar una descripción en la cual hay más de una opción.

[5] Sin embargo, lo característico es que haya varias estrategias para sobrevivir en el nivel de la subsistencia. En contraste con la elección del puente debajo del cual dormir, esas estrategias suelen diferir de manera no trivial. Uno de los defectos de la teoría del valor trabajo de Marx radica en que éste no entendió ese hecho.

[6] Otra idea del estructuralismo, no relacionada con ella, se examinó en el Capítulo 1.

[7] En otras palabras, paso por alto que la persona podría trabajar horas extra, elaborar su propia bebida alcohólica o comprar bienes de contrabando. En aplicaciones de políticas, estos problemas son de importancia.

[8] La forma de las curvas corresponde al hecho de que, cuanto más alcohol consume actualmente el agente, más necesita para compensar (y mantenerse en el mismo nivel de bienestar) un corte determinado en el lote de bienes de consumo.

[9] La sobrestimación del poderío militar soviético o iraquí podría recordarnos, sin embargo, que aun las oportunidades pueden ser difíciles de verificar.

[10] Aun cuando no tengamos razones para pensar que otra nación abriga hoy intenciones hostiles, pueden aparecer planes de contingencia para futuros escenarios. Es probable que en las profundidades del Departamento de Defensa de los Estados Unidos haya un plan para invadir Canadá.

[11] Además, como se argumenta más adelante, es posible que la mejor manera de modificar sus opiniones consista en modificar sus circunstancias.

[12] En contra de una creencia muy difundida, el fácil acceso de los oficiales de policía a las armas no los hace más propensos al suicidio que otros.

[13] En la década de 1970, al comprobarse una caída radical de los índices de suicidio en Gran Bretaña, el cambio se atribuyó en un comienzo a las líneas de ayuda establecidas por los Centros Samaritanos, pero más adelante se explicó sobre la base del reemplazo del letal gas de carbón de los hornos domésticos por el gas natural, menos peligroso en ese sentido.

[14] No creo que el aumento del *coste* de encontrar un medio pueda ser un factor disuasivo del suicidio. En la decisión de matarse hay tal vez consideraciones relacionadas con los costes y los beneficios, como la ponderación del dolor que el agente infligirá a otros en comparación con el alivio que él sentirá, pero el coste de encontrar un medio apropiado no representará obstáculo alguno para un individuo resuelto.

[15] De hecho, la mayoría de los estados norteamericanos imponen un tiempo de espera antes de la compra de un arma corta, cuyo fin es permitir a las autoridades constatar si el comprador potencial tiene un prontuario delictivo o un historial de enfermedad mental, y no interponer un plazo que «enfríe» al adquirente.

[16] Una incompatibilidad diferente se expresa en este dicho francés: *«Si jeunesse savait, si vieillesse pouvait*» (literalmente: si la juventud supiese, si la vejez pudiese), que da a entender que dos requisitos de la acción, el conocimiento de lo que es preciso hacer y la energía para hacerlo, nunca van juntos. Un proverbio noruego tiene un sesgo distinto: «Uno ha llegado a la madurez cuando es lo bastante viejo para saber bien que no debe y lo bastante joven para, aun así, hacerlo».

[17] La adaptación podría ser de importancia para el *bienestar* del agente, al aportarle paz espiritual. Este efecto, sin embargo, es ambiguo, porque el autoenvenenamiento de la mente asociado a muchos casos de «uvas verdes» o «limones dulces» (como los espectáculos de Broadway examinados en el Capítulo 1) puede, en una perspectiva más general, reducir el bienestar.

Capítulo 10

Personas y situaciones

La vergüenza y la culpa, o el desprecio y la ira, difieren en el hecho de que la primera emoción de cada par apunta al *carácter* de la persona y la segunda a alguna *acción* de ésta (Capítulo 8). De manera similar, la altivez se basa en la creencia de que uno es una persona superior, y el orgullo, en la creencia de que uno ha realizado algún hecho sobresaliente. Pero cuando censuramos o elogiamos una acción, ¿no lo hacemos porque creemos que refleja el carácter del agente? ¿A qué otro factor podría adjudicarse?

Cuando la psicología popular se equivoca

Este libro no se dedica al elogio o la censura, sino a la *explicación* del comportamiento. En ese contexto, la cuestión es la capacidad del carácter para explicar la acción. Con frecuencia se supone que la gente tiene rasgos de personalidad (introvertido, tímido, etc.), así como virtudes (honestidad, coraje, etc.) o vicios (los siete pecados capitales, etc.). En la psicología popular, esas características se presumen estables a lo largo del tiempo y a través de las situaciones. Proverbios de todas las lenguas dan testimonio de esa presunción. «El que miente una vez mentirá siempre», «el que miente también roba», «el que roba un huevo también robará una vaca», «el que cumple con su palabra en las pequeñas cosas, también lo hará en las grandes», «si te cogen con las manos en la masa, nunca confiarán en ti». Si la psicología popular tiene razón, la predicción y explicación del comportamiento debe ser cosa sencilla. Una sola acción revelará el rasgo o

la disposición subyacente y nos permitirá predecir el comportamiento en una cantidad indefinida de ocasiones en que la disposición pueda manifestarse. El procedimiento no es tautológico, como lo sería si consideráramos las trampas en un examen como prueba de deshonestidad y luego utilizáramos el rasgo de la deshonestidad para explicar las trampas. Equivale, en cambio, a utilizar las trampas en un examen como prueba de un rasgo (la deshonestidad) que también hará que la persona sea infiel a su cónyuge. Si uno acepta la teoría popular más extrema de que todas las virtudes van juntas, el trampear también podría usarse para predecir la cobardía en una batalla o el exceso en la bebida.

La gente suele hacer fuertes inferencias sobre la base de la austera conducta privada de otros. Se dice que un miembro de la Academia Francesa votó por De Gaulle debido a la dignidad de su vida privada, con la premisa tácita de que alguien que traicionara a su esposa también sería capaz de traicionar a su país. En Vietnam, los líderes comunistas pudieron conquistar «la mente y el corazón del pueblo» a causa de su incorruptible estilo personal, en marcado contraste con los organizadores menos abnegados de otros grupos políticos. Entre los mafiosos, se cree que la propensión a tener amoríos es el signo de un carácter desordenado y débil.

Hasta cierto punto, la psicología popular funciona como una profecía autocumplida. Si la gente *cree* que otros pronosticarán su comportamiento en una situación de tipo A sobre la base de su comportamiento en una situación de tipo B, actuará en esta última con la primera en mente. Si la creencia en un vínculo entre la moral privada y la moral pública tiene mucha difusión (y se sabe que la tiene), crea un incentivo para que los políticos se comporten honestamente en la vida privada, pues suponen que cualquier muestra de mala conducta llegará a conocimiento del electorado. Supongamos asimismo que disfruta de vasta aceptación la creencia de que las personas tienen el mismo índice de descuento temporal en todas las situaciones. Si se preocupan demasiado poco por su futuro para cuidar su cuerpo, también es probable (de acuerdo con la psicología popular) que rompan una promesa para obtener una ganancia a largo plazo. Por lo tanto, para poder hacer promesas creíbles en lo concerniente a una cooperación duradera y de beneficio mutuo, uno debe cultivar asimismo una apariencia esbelta y saludable.

Sin embargo, más allá de cierto punto *la psicología popular es demostrablemente falsa*.[18] Si se pueden eliminar sus efectos, de manera tal

que no haya incentivos para cumplir las expectativas de coherencia a través de las situaciones, se comprobará que esa coherencia es escasa. Los padres que sólo observan el comportamiento de sus hijos en casa suelen sorprenderse cuando se enteran de que se portan mucho mejor en la escuela o en lo de sus compañeros de clase. Por otra parte, las intervenciones para mejorar la conducta en la familia no llevan a un mejor ajuste en la escuela, en comparación con la situación de los grupos de control en los que no se ha intervenido.[19] En los experimentos de laboratorio, la mayoría de la gente (alrededor de dos tercios de los sujetos) puede ser inducida a comportarse sin piedad, al extremo de aplicar (lo que cree que son) grandes descargas eléctricas (de aproximadamente cuatrocientos cincuenta voltios) a un cómplice del experimentador. No hay, empero, razón para creer que su comportamiento se debe a un rasgo subyacente de sadismo, crueldad o indiferencia al sufrimiento de los otros; en realidad, muchos de los sujetos que se comportaron de ese modo se sentían perturbados y atormentados por lo que hacían. Los niños están mucho más dispuestos a esperar una recompensa demorada más grande cuando tanto ésta como otra más pequeña que podrían obtenerse de inmediato están ocultas a la vista. Todos los académicos conocen a colegas que son concienzudos en su trabajo, pero no lo son tanto en su actividad docente o administrativa. El hecho de ser locuaz en el almuerzo resulta tener una escasa correlación con la locuacidad en otras ocasiones. Una persona puede posponer una y otra vez la limpieza de la casa, pero no dilatar jamás las cosas en su trabajo.[20]

En uno de sus ensayos, «De la inconstancia de nuestros actos», Montaigne contrasta el comportamiento de Catón el Joven con el de seres humanos comunes y corrientes como él:

> Quien ha tocado una tecla [de Catón], ha tocado todas; es una armonía de sonidos que no puede desmentirse. Para nosotros, por el contrario, son precisos tantos juicios particulares como actos realizamos. A mi parecer, lo más seguro sería referirlos a las circunstancias próximas, sin entrar en más detalle ni sacar más conclusiones.

Como él mismo también señala: «Cuando siendo cobarde [un hombre] ante la infamia es firme ante la pobreza; cuando siendo débil rodeado por las navajas del barbero, resulta enérgico frente a las espadas de los adversarios, la acción es loable mas no así el hombre».

Querría citar algunos ejemplos del arte y los artistas. Proust escribió que podría haberse creído que los jóvenes a quienes en *Le Temps retrouvé* se paga para infligir dolor a los clientes del burdel de Jupien debían ser

«esencialmente malos, pero no sólo fueron en la guerra unos soldados maravillosos, unos incomparables "valientes", sino que, en muchos casos, fueron en la vida civil buenos corazones, si no unas excelentes personas». Al comentar la aparente contradicción entre la actitud de Swann al «disimular con tanta gracia una invitación al Buckingham Palace» y su jactancia al decir que la mujer de un funcionario subalterno había visitado a la señora Swann, señala:

> la razón principal, razón que puede aplicarse a la humanidad en general, es que ni siquiera nuestras virtudes son cosa libre y flotante, cuya permanente disponibilidad conservamos siempre, sino que acaban por asociarse tan estrechamente en nuestro ánimo con las acciones que nos imponen el deber de ejercitar las dichas virtudes, que si surge para nosotros una actividad de distinto orden nos encuentra desprevenidos y sin que se nos ocurra siquiera que esta actividad podría traer consigo el ejercicio de esas mismas virtudes.

Un médico que lo conocía caracterizó a Charlie Parker, el músico de *jazz*, como «un hombre que vive en el momento. Un hombre que vive para el principio de placer, la música, la comida, el sexo, las drogas, la diversión, con su personalidad [*sic*] estancada en un nivel infantil». Otro gran músico de *jazz*, Django Reinhardt, tenía una actitud imprevisora aún más extrema en su vida diaria: nunca ahorraba un centavo de sus abundantes ingresos y los gastaba todos en caprichos o autos costosos, que rápidamente procedía a chocar. En muchos aspectos, era la encarnación del estereotipo del «gitano». Sin embargo, uno no se convierte en un músico del calibre de Parker y Reinhardt si vive en el momento *en todos los sentidos*. La excelencia implica años de dedicación y concentración absolutas. En el caso de Reinhardt, esa dedicación salió dramáticamente a la luz cuando sufrió un grave daño en la mano izquierda a raíz de un incendio y decidió reejercitarse para estar en condiciones de hacer más con dos dedos que cualquier otro con cuatro. Si estos dos músicos hubieran sido impulsivos y despreocupados en todos los ámbitos (si su «personalidad» hubiese sido uniformemente «infantil»), jamás podrían haber llegado a ser artistas tan consumados.

Luego de 1945, el novelista noruego Knut Hamsun, que había colaborado con los nazis durante la guerra, fue sometido a una observación psiquiátrica para determinar si su estado mental permitía llevarlo a juicio (tenía por entonces ochenta y seis años). Cuando el profesor de psiquiatría que lo examinaba le pidió que describiera sus «principales rasgos de carácter», respondió lo siguiente:

> Los escritores del llamado período naturalista, Zola y su tiempo, sólo se referían a los rasgos principales de carácter de las personas. No veían utilidad en los matices de la psicología. La gente tenía una capacidad dominante que gobernaba sus actos. Dostoievski y otros nos enseñaron algo diferente sobre el hombre. No creo que, desde el comienzo mismo, haya en ninguno de mis escritos una sola persona con esta capacidad dominante y unitaria. Todos carecen del llamado carácter: están divididos y fragmentados y no son ni buenos ni malos, sino ambas cosas a la vez. Variopintos y cambiantes en su mente y sus actos. Es indudable que yo mismo soy así. Es muy posible que sea agresivo y que tenga algunos de los demás rasgos sugeridos por el profesor: vulnerable, receloso, egoísta, generoso, celoso, probo, lógico, sensible, de talante frío. Todos estos serían rasgos humanos, pero no puedo atribuir a ninguno de ellos preponderancia en mi persona.

En el Capítulo 14 abordaré la cuestión del carácter o la «falta de carácter» en las obras de ficción. Aquí me limitaré a señalar que Hamsun no se refiere a la posibilidad de ser, por ejemplo, uniformemente generoso en un tipo de situación y uniformemente egoísta en otro. Me ocupo ahora de este problema.

El poder de la situación

«Ser impulsivo con el dinero» y «estar dedicado a su música», «ser conversador en el almuerzo» o «ser concienzudo en sus investigaciones» también son, desde luego, rasgos de carácter. Se trata, no obstante, de *rasgos locales* o específicos de situaciones, y no de características generales de la personalidad que se manifiesten en todos los ámbitos y todas las circunstancias. Contra lo que supone la psicología popular, los estudios sistemáticos encuentran niveles muy bajos de coherencia de los rasgos de carácter a través de las situaciones. Si bien hay correlaciones, por lo común son tan bajas que no pueden detectarse «a simple vista». Los psicópatas quizás exhiban de manera general un comportamiento insensible[21] y Catón el Joven puede haber sido coherentemente heroico, pero en el caso de la gran mayoría de los individuos que se encuentran entre estos dos extremos no debe esperarse esa coherencia. Una idea más extrema de la psicología popular, la de que todas las virtudes van juntas, no ha sido objeto de exámenes tan exhaustivos, tal vez porque su inverosimilitud parece obvia. No obstante, puede tener cierto influjo sobre la mente, como lo muestra nuestra confianza en las destrezas profesionales de los médicos que tienen buenos «modales de cabecera». En la antigüedad clásica, la idea de que la excelencia en un ámbito era un predictor o «indicio» infalible de la

excelencia en otros era habitual. Cuando se refieren a ella, los psicólogos hablan de «efecto halo».

En consecuencia, la explicación del comportamiento se encuentra a menudo en la *situación* y no en la *persona*. Consideremos, por ejemplo, el hecho de que algunos alemanes hayan actuado para rescatar a judíos del régimen nazi. Conforme a una teoría «caracterológica», cabría suponer que esas personas tenían un tipo de personalidad altruista de la que carecían quienes no rescataron a nadie. Sin embargo, resulta que el factor con mayor capacidad explicativa era el hecho «situacional» de que se *les pidiera* rescatar a alguien. El vínculo causal podía surgir de dos maneras. Por un lado, sólo al recibir el pedido puede uno obtener la *información* de que es necesario actuar como rescatador. Por otro, la situación cara a cara en que se solicita la ayuda podría provocar la aceptación, debido a la *vergüenza* que uno sentiría si se negara.[22] La primera explicación supone el altruismo pero niega que sea suficiente para explicar el comportamiento. La segunda niega el altruismo y sustituye las normas morales por normas sociales. De una u otra manera, lo que diferencia a los rescatadores de los no rescatadores es la situación en que se encuentran, y no su personalidad.

El caso de Kitty Genovese es otro ejemplo de la vida real del poder de la situación. Es poco verosímil estipular, sobre la base de su inacción, que todos los testigos de su asesinato eran insensibles e indiferentes al sufrimiento humano. Antes bien, es posible que muchos de ellos hayan pensado que otra persona iba a llamar a la policía, que, como nadie hacía nada al respecto, la situación no era tan grave como parecía («probablemente sólo una riña doméstica»), o que la inacción de los demás sugería que la intervención directa podía ser riesgosa.[23] Estas líneas de razonamiento resultan más plausibles cuanto mayor es la cantidad de espectadores. Así, en un experimento, los sujetos escucharon a un cómplice del experimentador fingir un ataque de epilepsia a través del sistema de intercomunicación. Cuando cada sujeto creía que sólo él escuchaba, la intervención para ayudar ascendió al 85%; cuando creían que había otro, intervino el 62%, y cuando creían que el total de oyentes era cinco, esa proporción cayó al 31%. En otro experimento, intervino el 70% de los espectadores solos, pero apenas el 7% lo hizo cuando a su lado se sentaba un compañero impasible. Con dos sujetos ingenuos, la víctima recibió ayuda en el 40% de los casos. De tal modo, no sólo cae la posibilidad de que cualquier espectador *dado* intervenga cuando hay otros involucrados;

también disminuye la posibilidad de que *alguno* lo haga al incrementarse el número de espectadores.[24] En otras palabras, la dilución de la responsabilidad de intervenir causada por la presencia de los otros se produce de manera tan acelerada que el mayor número de potenciales intervinientes no puede compensarla.

En otro experimento, se indicó a unos estudiantes de teología que se prepararan para dar una breve charla en un edificio cercano. La mitad recibió la instrucción de basar la charla en la parábola del buen samaritano (!), mientras que el resto debía hablar de un tema más neutro. Se dijo a un grupo que debía apresurarse, dado que las personas del otro edificio ya estaban esperando, mientras que al otro se le informó que tenía mucho tiempo. En su camino al otro edificio, los sujetos tropezaron con un hombre que se desplomaba junto a una puerta, aparentemente en apuros. Entre los estudiantes que presuntamente estaban atrasados, sólo el 10% se ofreció a ayudarlo; en el otro grupo lo hizo el 63%. El grupo que tenía la instrucción de preparar una charla sobre el buen samaritano no mostró más inclinación a comportarse como tal. El comportamiento de los estudiantes tampoco se correlacionó con las respuestas a un cuestionario destinado a evaluar si su interés en la religión se debía al deseo de salvación personal o a un deseo de ayudar a otros.[25] El factor situacional, estar apresurados o no, tenía mucha más capacidad explicativa que cualquier factor disposicional.

No sería exacto subsumir este análisis en el análisis del capítulo anterior y decir que los estudiantes de la categoría «con prisa» se comportaron como lo hicieron debido a una *restricción temporal*. Su restricción no era objetiva ni «dura», y de hecho el 10% de los integrantes de este grupo ofreció su asistencia. Antes bien, fue la situación la que dio forma al comportamiento al afectar la preponderancia de *deseos* antagónicos. El pedido cara a cara incrementa la fuerza de los motivos referidos a otros, mientras que la indicación de apresurarse la reduce. La capacidad de *ver* la recompensa cuya disponibilidad es inminente la hace más atractiva si se la compara con otra que sólo estará a nuestro alcance al cabo de un tiempo, así como la visión de un mendigo en la calle puede despertar una generosidad que el conocimiento abstracto de la pobreza no suscitaría. Las situaciones del tipo de Kitty Genovese modifican la percepción tanto de los costes como de los beneficios de la ayuda. El deseo de cumplir la instrucción de un experimentador impasible en el sentido de que «es preciso continuar» con la aplicación de descargas eléctricas aparentemente dolorosas y posiblemente

fatales pasa por encima del deseo de no infligir dolor de manera innecesaria.

No hay un mecanismo general o común mediante el cual una situación pueda afectar el comportamiento. Las situaciones varían desde los pedidos cara a cara para colaborar en el rescate de judíos hasta los sucesos más triviales, como cuando el hallazgo de una moneda de veinticinco centavos en la ranura de un teléfono público nos levanta el ánimo y nos lleva a ayudar a un extraño (en realidad un cómplice del experimentador) a recuperar un manojo de papeles caídos en la acera. La lección de importancia que debe extraerse de estas observaciones, en la vida real y en el laboratorio, es simplemente que, *con frecuencia, el comportamiento no es más estable que las situaciones que le dan forma*. Una persona tal vez sea conversadora en el almuerzo, cuando puede relajarse en compañía de viejos colegas, y se le trabe la lengua en presencia de extraños. Otra quizá dé siempre algo a los mendigos, sin pensar, por otra parte, una sola vez en los pobres. Una persona puede prestar invariablemente ayuda en situaciones en las que nadie más puede hacerlo, y ser invariablemente pasiva en presencia de otros potenciales ayudantes. Un hombre tal vez se muestre constantemente agresivo con su mujer y le haga mordaces observaciones, pero sea calmo y generoso con otras personas. Su mujer también podría exhibir el mismo comportamiento doble. La agresión de él desata la de ella, y a la inversa.[26] Si cada uno de ellos ve contadas veces al otro interactuar con otros adultos, por ejemplo en el lugar de trabajo, quizá se consideren en forma recíproca intrínsecamente agresivos, y no sólo agresivos en la situación definida por su presencia.

La apelación espontánea a las disposiciones

Para seguir con el último ejemplo, los terapeutas de pareja procuran, con frecuencia, hacer que los cónyuges que han buscado su ayuda pasen de un lenguaje basado en el carácter a un lenguaje basado en la acción. En vez de decir «eres una mala persona», y dar con ello poca cabida a la esperanza o el cambio, deberían hacer el esfuerzo de decir «has hecho algo malo». Esta última formulación deja abierta la posibilidad de que la acción en cuestión haya sido desencadenada por factores situacionales específicos, como una observación provocativa del otro miembro de la pareja. Una razón (entre muchas) por la cual los terapeutas suelen tener poco éxito

cuando reformulan de ese modo los conflictos, es que la gente privilegia en forma espontánea las explicaciones del comportamiento basadas en el carácter en desmedro de las basadas en la situación. Si nos enteramos de que alguien ha contribuido a un aviso sobre los «derechos de los *gays*», tendemos a suponer que esa persona *es gay* o liberal, en vez de pensar que acaso *le hayan pedido* su colaboración de un modo tal que le resultaba difícil negarse. Cuando realizamos una entrevista laboral a una persona, tendemos a explicar lo que ésta dice o hace a la luz de las disposiciones que (con exceso de seguridad) le atribuimos, en vez de adjudicarlo a la naturaleza especial de la situación de entrevista. El lenguaje mismo refleja este sesgo disposicional. Los adjetivos que se aplican a las acciones («hostil», «egoísta» o «agresiva») también pueden aplicarse por lo común al agente, mientras que son pocas las aplicables asimismo a las situaciones («difícil» es una excepción).

Para referirse al uso inapropiado de la explicación disposicional, los psicólogos hablan de *error fundamental de atribución*, esto es, el hecho de explicar un comportamiento inducido por la situación como si lo hubiesen causado rasgos de carácter perdurables del agente. Cuando se les pidió que predijeran el comportamiento de los estudiantes de teología que tropezaron con un individuo en apuros, los sujetos de un experimento creyeron (equivocadamente) que personas cuya religión se fundaba en un deseo de ayudar a otros se inclinarían más a actuar como buenos samaritanos y (otra equivocación) que la mayor o menor prisa no influiría en absoluto. Otros sujetos predijeron en exceso la aplicación de descargas eléctricas en ausencia de factores situacionales específicos en el experimento original, y revelaron de tal modo su creencia en una explicación disposicional. Cuando un instructor asigna a un estudiante la tarea de escribir un texto a favor de Castro, otros estudiantes, pese a saber cómo fue asignada, consideran de todos modos que el escrito expresa una actitud procastrista. Cuando se pidió a un grupo de estudiantes que se ofrecieran para realizar tareas que tenían una remuneración alta o baja, y como resultado hubo cantidades grandes o pequeñas de voluntarios, los observadores, si bien conocían el diferencial de pago, predijeron no obstante que *todos* los voluntarios se inclinarían más que los no voluntarios a ofrecerse a trabajar por una causa no remunerada. En otras palabras, los observadores atribuían la acción de ofrecerse como voluntarios a una disposición a hacerlo, y no a la estructura de recompensas de la situación.

Los integrantes de algunas sociedades parecen menos propensos que los de otras al error fundamental de atribución. Los experimentos indican que, en comparación con los estadounidenses, los asiáticos adjudican mayor importancia a la situación y menor importancia a las disposiciones personales cuando se trata de explicar un comportamiento. También las situaciones de la vida real exhiben esta diferencia. Así, en 1991, un frustrado estudiante chino de física disparó contra su consejero y varios de sus compañeros y luego se mató. Ese mismo año, un empleado postal estadounidense que había perdido su trabajo atacó a tiros a su supervisor y a varios compañeros y transeúntes y después se suicidó. Ambos sucesos tuvieron mucha difusión en diarios estadounidenses y chinos; los primeros los explicaban de manera uniforme en términos disposicionales («perturbado», «mal genio», «inestabilidad mental»), y los segundos, en términos situacionales («fácil acceso a las armas», «acababa de quedar cesante», «víctima de la presión por tener éxito»). Otros descubrimientos confirman esta diferencia, que podría deberse, no obstante, al hecho de que los factores situacionales cumplen en realidad un papel más grande en la generación de comportamientos entre los asiáticos. En vez de ser mejores para superar el sesgo disposicional, éstos tal vez disten de tener que superar un sesgo; otra posibilidad es que actúen ambos tipos de factores.

La superación del error fundamental de atribución puede ser liberadora. Los estudiantes universitarios de primer año a quienes se dice que la mayoría de los recién ingresados tienen un mal desempeño, pero que más adelante sus calificaciones mejoran, de hecho se desempeñan mejor en años ulteriores que aquellos a quienes no se da esa información. Es más probable que estos últimos atribuyan su pobre desempeño a su escasa aptitud y no a un entorno universitario poco conocido y lleno de distracciones. Al no creer que les pueda ir mejor, están menos motivados para tratar de hacerlo. Cuando los grupos oprimidos se deshacen del esencialismo de sus opresores –la idea de que las mujeres, los negros o los judíos son intrínsecamente inferiores–, les resulta más fácil liberarse de sus cadenas.

¿El error fundamental de atribución es caliente o frío: una equivocación motivada o algo de naturaleza más afín a una ilusión óptica? En la medida en que la motivación interviene en el proceso de atribución, no hay razón por la cual deba inducirnos de manera uniforme a hacer excesivo hincapié en las disposiciones. En un terreno egocéntrico, deberíamos atribuir nuestro éxito a nuestros rasgos de carácter perdurables, y nuestros fracasos a

circunstancias desafortunadas.[27] Y de dar crédito a los moralistas franceses, debemos atribuir los éxitos de otros a su buena suerte y sus fracasos a sus disposiciones.[28] En un terreno cognitivo, la tendencia a favorecer a la persona en detrimento de la situación puede ser un ejemplo de una tendencia más general a prestar mayor atención al primer plano en movimiento que al segundo plano estático. De ello se deduce que el error debería ser menos habitual en las culturas en que se presta una atención más igualitaria a ambos planos, como parece suceder entre los asiáticos.

La rehabilitación de la persona

Los descubrimientos que he descrito socavan lo que podríamos llamar «esencialismo crudo» en el estudio de la personalidad. Sencillamente, no es cierto que las personas *sean* agresivas, impacientes, extrovertidas o conversadoras en todo momento y lugar. Al mismo tiempo, dichos hallazgos no dan a entender que la situación sea todopoderosa para explicar el comportamiento. Antes bien, tenemos que descomponer «el» carácter en un conjunto de tendencias a las respuestas *contingentes*. En vez de caracterizar a una persona como altruista, podríamos describirla con las frases «ayuda cuando se lo piden, pero no se ofrece voluntariamente para hacerlo» o «ayuda cuando no está estresada, pero es negligente cuando lo está». Cada una de esas frases podría caracterizar un aspecto de la persona y, de tal modo, suscribir una forma más sutil de esencialismo. Una persona quizá riña a su cónyuge por no limpiar nunca la casa («eres perezoso») o por no hacerlo nunca a menos que se lo pidan («eres desconsiderado»). En este último caso, el cónyuge podría ser proactivo y no reactivo en otros asuntos, como el control de la salud de los hijos. No habría un rasgo general de reactividad.

Desde este punto de vista, la explicación del comportamiento se apoya en la situación particular más la relación específica de la persona entre las situaciones y la propensión de conducta. Una persona podría ser en extremo agresiva con individuos sobre los cuales tiene poder, pero excepcionalmente amigable con aquellos que tienen poder sobre ella, mientras que otra podría exhibir el patrón opuesto. Si vemos a las dos comportarse de una manera amigable, quizá sintamos la tentación de concluir que ambas *son* de una disposición amistosa. Sin embargo, como ya tendría que resultar claro, la similitud de los comportamientos puede deberse a diferencias en la

situación y en contingencias de respuesta que se neutralizan exactamente unas a otras.

* * *

NOTA BIBLIOGRÁFICA

La «tendencia a sobrestimar la unidad de la personalidad» fue expuesta con claridad por Gustav Ichheiser, «Misunderstandings in human relations: a study in false social perception», *American Journal of Sociology*, 55(2), septiembre de 1949, suplemento, págs. 1-72. Los trabajos recientes que reducen el papel del «carácter» derivan de Walter Mischel, *Personality and Assessment*, Nueva York, Wiley, 1968 [*Personalidad y evaluación*, México, Trillas, 1973]. La presente exposición se apoya mucho en Lee Ross y Richard E. Nisbett, *The Person and the Situation: Perspectives of Social Psychology*, Filadelfia, Temple University Press, 1991, y en John M. Doris, *Lack of Character: Personality and Moral Behavior*, Cambridge y Nueva York, Cambridge University Press, 2002. Las referencias a los organizadores comunistas en Vietnam y a los mafiosos provienen, respectivamente, de Samuel Popkin, *The Rational Peasant: The Political Economy of Rural Society in Vietnam*, Berkeley, University of California Press, 1979, y Diego Gambetta, «Trust's odd ways», en Jon Elster y otros (comps.), *Understanding Choice, Explaining Behavior: Essays in Honour of Ole-Jørgen Skog*, Oslo, Oslo Academic Press, 2006, págs. 81-100. La referencia a los efectos de la intervención corresponde a Judith Rich Harris, *No Two Alike*, Nueva York, Norton, 2006. La disposición a aplicar descargas eléctricas se describe en un estudio clásico de Stanley Milgram, *Obedience to Authority: An Experimental View*, Nueva York, Harper, 1983 [*Obediencia a la autoridad: un punto de vista experimental*, Bilbao, Desclée de Brouwer, 1980]. Se encontrará información sobre los dos músicos mencionados en Ross Russell, *Bird Lives: The High Life and Hard Times of Charlie (Yardbird) Parker*, Nueva York, Charterhouse, 1973 [*Bird: Biografía de Charlie Parker*, Barcelona, Ediciones B, 1989], y Michael Dregni, *Django: The Life and Music of a Gypsy Legend*, Oxford y Nueva York, Oxford University Press, 2004. La declaración de Hamsun se ha traducido de Gabriel Langfeldt y Ømulv Ødegård, *Den rettpsykiatriske erklæringen om Knut Hamsun*, Oslo, Gyldendal, 1978, pág. 82. El uso del comportamiento como un «indicio» en la antigüedad es analizado por Paul

Veyne, *Le Pain et le cirque: sociologie historique d'un pluralisme politique*, París, Seuil, 1976, págs. 114 y 773; véase también Paul Veyne, «Pourquoi veut-on qu'un prince ait des vertus privées? Ou de la nécessité de savoir à quoi on a affaire», *Social Science Information*, 37(3), 1998, págs. 407-415. La explicación «caracterológica» de la voluntad de rescatar judíos es sostenida por Kristen Monroe, Michael C. Barton y Ute Klingemann, «Altruism and the theory of rational action: rescuers of Jews in Nazi Europe», *Ethics*, 101, octubre de 1990, págs. 103-122. La explicación «situacionista» se expone en Federico Varese y Meir Yaish, «The importance of being asked: the rescue of Jews in Nazi Europe», *Rationality and Society*, 12, 2000, págs. 307-324. En un análisis del caso de Kitty Genovese basado en la teoría de los juegos, Avinash Dixit y Susan Skeath, *Games of Strategy*, segunda edición, Nueva York, Norton, 2004, págs. 414-418, argumentan que debemos esperar la reducción de las posibilidades de que *alguien* intervenga cuando el número de intervinientes potenciales aumenta. Una nota escéptica acerca de la tendencia a inferir disposiciones del comportamiento resuena en James L. Hilton, Steven Fein y Dale Miller, «Suspicion and dispositional inference», *Personality and Social Psychology Bulletin*, 19(5), 1993, págs. 501-512. El contraste entre estadounidenses y asiáticos se resume en Richard E. Nisbett, *The Geography of Thought: How Asians and Westerners Think Differently –and Why*, Nueva York, Free Press, 2004. Lo que denomino «rehabilitación de la persona» se plantea en Walter Mischel, «Toward an integrative science of the person», *Annual Review of Psychology*, 55, febrero de 2004, págs. 1-22.

Notas al pie

[18] No sólo la psicología popular: los especialistas en ciencias sociales que sostienen que, por medio de su comportamiento, los agentes señalan si son «buenos tipos» o «malos tipos» también sobrestiman la consistencia.

[19] Estos descubrimientos tienen un truco. Los hijos de los padres que siguieron el consejo de los organizadores de la intervención funcionaron mejor en la escuela que los hijos de quienes no lo siguieron, un hecho mencionado por algunos en respaldo del derrame del hogar a la escuela. Sin embargo, este descubrimiento puede deberse simplemente al hecho de que el cumplimiento es heredable. Los padres que siguen de manera concienzuda las instrucciones de un intervencionista tienen mayores probabilidades de tener hijos que sigan de manera concienzuda las instrucciones de un docente.

[20] En una carta al eticista Randy Cohen (*The New York Times Magazine*, 15 de enero de 2006), un académico le preguntó si el hecho de que un colega no numerario reclamara en el Club de Profesores unos descuentos a los que no tenía derecho, justificaba un voto en su contra en la elección para un cargo, «debido a su deshonestidad y la extensión potencial de ésta a su investigaciones». Cohen dijo que no, con el argumento de que «las personas que se comportan mal en algunas situaciones a menudo se comportan bien en otras».

[21] Como un egoísta inteligente que se preocupara por el futuro del mundo tendría con frecuencia interés en *remedar* la preocupación por los otros (Capítulo 5), la explicación última de la conducta psicopática podría radicar en un descuento excesivo del futuro.

[22] De manera similar, el éxito de las campañas televisivas de beneficencia que logran inducir a la gente a donar dinero no se basa en su apelación a motivos altruistas, sino al hecho de que las acompaña un llamado a la puerta en que alguien hace un pedido cara a cara. En este caso, la explicación basada en la información es claramente inadecuada.

[23] Las personas que tenían miedo de intervenir materialmente para proteger a la víctima de su atacante habrían podido, de todos modos, llamar a la Policía. En esa época, sin embargo, ésta no aceptaba llamados anónimos, de modo que los espectadores del hecho quizá temieran meterse en líos. En otras situaciones de este tipo, puede no existir la opción de hacer ese llamado.

[24] Ésa es, al menos, la tendencia en los numerosos experimentos de este tipo que se han llevado a cabo. En el ejemplo recién citado (el ataque de epilepsia escuchado a través del intercomunicador), resulta que, si suponemos que los otros escuchas eran sujetos reales ingenuos que habían recibido la misma información (y no meros cómplices o ficciones creadas por el experimentador), la posibilidad de que al menos uno de ellos interviniera es aproximadamente constante, en torno del 85%. En el caso de cinco sujetos (el sujeto principal y los cuatro escuchas), la posibilidad de que alguno de ellos se abstuviera de intervenir es de 0,69. La posibilidad de que todos se abstuvieran es de $(0,69)^5$ o 0,156, lo que da una probabilidad de 0,844 de que al menos uno interviniese.

[25] Al igual que los sujetos a quienes se indujo a aplicar descargas eléctricas, muchos de los que pasaron a toda prisa junto al hombre en apuros estaban visiblemente afligidos por el encuentro.

[26] Metafóricamente hablando, se encuentran en un «mal equilibrio psicológico». Sin embargo, la agresión no tiene por qué ser la «mejor respuesta» a la agresión (según lo exige una noción del equilibrio basada en la teoría de los juegos); sólo es una respuesta psicológicamente inteligible.

[27] A veces, sin embargo, tal vez estemos motivados a adjudicar nuestros fracasos a nuestro carácter. Un jugador o un alcohólico quizá se sientan satisfechos al decirse que «no pueden evitarlo», a fin de tener una excusa para persistir.

[28] En ocasiones, sin embargo, tal vez estemos motivados a atribuir los éxitos de otros a sus *defectos* de carácter. El antisemitismo se apoya en el mito de que los judíos son exitosos porque, debido a su carácter inmoral, están dispuestos a hacer cualquier cosa con tal de prosperar.

Capítulo 11

Elección racional

La estructura de la acción racional

Los teóricos de la elección racional aspiran a explicar el comportamiento sobre la mera base del supuesto de que los agentes son racionales. Este supuesto incluye la hipótesis de que los agentes se forman creencias racionales, entre ellas las referidas a las opciones que tienen a su alcance. No es necesario, por tanto, clasificar los determinantes del comportamiento sea como subjetivos (deseos) o como objetivos (oportunidades). La teoría de la elección racional es subjetiva de cabo a rabo.

La estructura de explicación basada en la elección racional se expone en la Figura 11.1. De conformidad con este esquema, una acción es racional si satisface *tres requisitos de optimalidad*: la acción debe ser óptima, dadas las creencias; las creencias deben tener el mejor respaldo posible, dadas las pruebas; y las pruebas deben ser la resultante de una inversión óptima en la recolección de información. En la Figura 11.1, las flechas tienen una doble interpretación, en términos tanto de causalidad como de optimalidad. La acción, por ejemplo, debería ser causada por los deseos y creencias que hacen que sea racional; no basta hacer lo correcto por azar. De manera similar, una creencia no es racional si es el resultado de dos procesos sesgados en direcciones opuestas que se neutralizan exactamente uno a otro. Para tomar un ejemplo, tanto fumadores como no fumadores procesan la información acerca de los peligros del hábito de fumar de una manera que

los lleva a creer que éstos son más grandes de lo que en realidad son. Al mismo tiempo, los fumadores están sujetos a un sesgo egocéntrico que les hace descontar los riesgos. Si como consecuencia de ello se forman la misma creencia que sostendría un observador imparcial,[29] esto no prueba que sean racionales. En uno de los análisis más importantes de la racionalidad en las ciencias sociales, Max Weber cometió el error de inferir la «racionalidad de los procesos» de la «optimalidad de los resultados», cuando escribió que

> a los efectos de un análisis tipológico científico, es conveniente tratar todos los elementos irracionales y afectivamente determinados del comportamiento como factores de desviación de un tipo conceptualmente puro de acción racional. Por ejemplo, la manera más adecuada de analizar un pánico bursátil consiste en intentar determinar ante todo cuál habría sido el curso de acción de no haber estado influenciado por afectos irracionales; es posible introducir luego los componentes irracionales como explicación de las desviaciones observadas de ese curso hipotético. De manera similar, al analizar una campaña política o militar es conveniente determinar en primer lugar cuál habría sido un curso racional, dados los fines de los participantes y un conocimiento apropiado de todas las circunstancias. *Sólo de este modo* es posible evaluar la significación causal de los factores irracionales como explicación de la desviación con respecto a ese tipo.

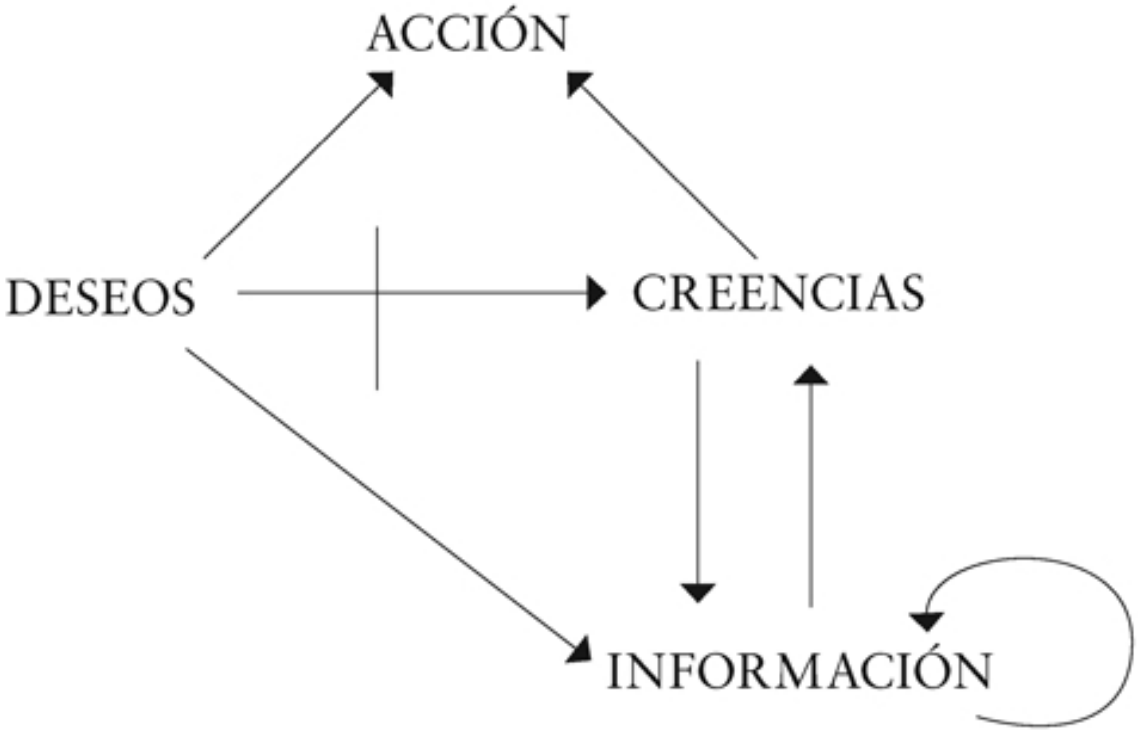


FIGURA 11.1

Aunque Weber acertaba al considerar que la desviación con respecto al curso de acción racional es una condición suficiente para señalar la intervención de la irracionalidad, se equivocaba al aseverar (en la frase que he puesto en bastardillas) que era una condición necesaria. Un error similar se comete al afirmar que las reacciones instintivas de miedo son racionales, cuando todo lo que puede decirse es que son *adaptativas*. Cuando veo en el sendero una forma que puede ser o bien un palo o bien una serpiente, es sensato salir corriendo de inmediato en vez de recoger más información. Al parecer, los seres humanos están de hecho predeterminados para actuar de ese modo. Este comportamiento de huida no es racional en sentido estricto, dado que no es un producto de la maquinaria de la toma de decisiones racionales, no obstante lo cual remeda la racionalidad en cuanto es el mismísimo comportamiento que dicha maquinaria habría producido de haberse aplicado a la situación. Cuando los costes de oportunidad de la recolección de información (Capítulo 12) son elevados, un agente racional no recogerá mucha. Sin embargo, a menudo la tendencia a la huida no es la resultante de esos cálculos, sino que se les anticipa.[30]

Preferencias y utilidad ordinal

Explicado con mayor detalle, el primer requisito de optimalidad es que la acción debe ser el mejor medio de satisfacer los deseos del agente, dadas sus creencias sobre las opciones disponibles y sus consecuencias. Lo «mejor», en el sentido de «lo óptimo», se define en términos de «perfección» o preferencia: aquello que, a juicio del agente, es mejor que todo lo demás. No hay implicación alguna de que el deseo sea *egoísta*. La confusión de la racionalidad y el egoísmo es un grosero error, aunque la práctica de algunos teóricos de la elección racional facilita cometerlo. Tampoco es necesario estipular que los deseos son *estables*, ni siquiera en el sentido mínimo de excluir los cambios temporarios de preferencias. Un agente que, bajo la influencia de la emoción o las drogas, prefiere A a B, actúa de manera racional al escoger A, aun cuando en otras circunstancias prefiera B a A. Viene a cuento lo que ocurre (véase el Capítulo 6) cuando el valor que el agente atribuye a las consecuencias futuras de la elección presente queda disminuido como resultado de aquellas influencias.

Para que el análisis levante vuelo, es preciso definir bien la noción de «lo óptimo». Dos condiciones aseguran que así sea.[31] En primer lugar, las

preferencias tienen que ser *transitivas*. Supongamos que hay tres opciones, A, B y C. Si una persona piensa que A es al menos tan buena como B y B al menos tan buena como C, también debe pensar que A es al menos tan buena como C. Si no hay transitividad, por ejemplo si la persona prefiere estrictamente A a B, B a C y C a A, es posible que no tenga una opción «óptima». Además, otra persona puede aprovechar ese hecho y ofrecer al agente pasar de una opción con menor preferencia a otra con mayor preferencia a cambio de una suma de dinero. Como las preferencias pasan por ciclos, esta operación puede repetirse indefinidamente y provocar la ruina de la persona mediante una serie de mejoras escalonadas.[32]

Esta situación puede darse si el agente clasifica las opciones por los «aspectos que cuentan». Supongamos que prefiero una manzana a otra si la primera es mejor en al menos dos de tres aspectos, como el precio, el sabor y la durabilidad. Si la manzana A supera a la manzana B en precio y sabor, la manzana B supera a la manzana C en precio y durabilidad y la manzana C supera a la manzana A en sabor y durabilidad, la transitividad queda invalidada. Aunque esta posibilidad es relativamente poco importante en la elección individual, en la cual no hace sino reflejar el fracaso de una regla práctica, veremos (Capítulo 25) que es más significativa en la elección colectiva.

Un problema diferente se plantea cuando la indiferencia no es transitiva. Me pueden dar lo mismo A y B y B y C, porque las diferencias dentro de cada par son demasiado pequeñas para advertirlas, pero prefiero C a A porque hay una diferencia detectable entre ellas. Hay una opción que es «óptima», a saber, C, pero pese a ello es posible empeorar la situación del agente si se le hace una serie de ofrecimientos, cambiar C por B y B por A, que no tiene motivos para rechazar y, por ende, bien podría aceptar. Lo que justifica calificar de irracional a un agente con preferencias intransitivas no es tanto la falta de una opción «óptima», sino el hecho de que puede aceptar ofrecimientos que lo hacen estar peor.

Para garantizar que la idea de «lo óptimo» siempre sea valedera, también debe requerirse que las preferencias sean *completas*: para dos resultados cualquiera, el agente debe poder decir si prefiere el primero al segundo o el segundo al primero o es indiferente entre ambos. Si no puede dar ninguna de esas tres respuestas, tal vez no esté en condiciones de determinar cuál de las opciones es la mejor. Diré más sobre la incompletitud hacia el final del capítulo. Aquí, sólo quiero señalar que, a diferencia de la

falta de transitividad, una falta de completitud no significa ningún tipo de fracaso. Supongamos que, entre dos niños, quiero dar un helado al que más ha de disfrutarlo. Para que *yo* tenga una preferencia con respecto a las dos opciones, tendría que poder comparar los niveles de satisfacción de preferencias *de ellos* en caso de que les diera el helado. A menudo, sin embargo, la tarea es imposible. No lograr llevarla a cabo no es un fracaso, en el sentido de haber podido hacer mejor las cosas, y refleja simplemente un hecho de la vida.

En muchos aspectos, la transitividad y la completitud de las preferencias son todo lo que necesitamos para identificar la acción racional. Sin embargo, con frecuencia es conveniente representar las preferencias mediante números, muchas veces llamados *valores de utilidad*, que se asignan a las opciones. Para asegurarnos esta posibilidad, imponemos una nueva condición a las preferencias: la *continuidad*. Si cada opción en una secuencia A1, A2, A3, …, se prefiere a B y la secuencia converge en A, A debería preferirse a B; si B se prefiere a cada opción de la secuencia, B debería preferirse a A. Las «preferencias lexicográficas» proponen un contraejemplo: un grupo de dos bienes A y B en cantidades (A1, B1) se prefiere a otro grupo (A2, B2) si y sólo si o bien A1 > A2 ó (A1 = A2 y B1 > B2). En esta clasificación de preferencias, los grupos (1,1, 1), (1,01, 1), (1,001, 1), …, se prefieren a (1, 2), que a su vez se prefiere a (1, 1). En términos aproximados, podemos decir que el primer componente del grupo es incomparablemente más importante que el segundo, dado que ninguna cantidad extra del bien B puede compensar siquiera la más mínima pérdida del bien A.[33] O, más sencillamente, ninguna solución de compromiso es posible. Por consiguiente, estas preferencias no pueden representarse mediante curvas de indiferencia. Si bien las preferencias lexicográficas no se aplican nunca o casi nunca a los bienes corrientes de consumo, pueden ser de importancia para las elecciones políticas. Un votante quizá prefiera al candidato A antes que al candidato B si y sólo si A tiene una postura más vigorosa en contra del aborto *o* si ambos tienen la misma postura en esa cuestión y A propone impuestos más bajos que B. Para esos votantes, el «valor sagrado» de la vida no puede intercambiarse por el valor secular del dinero.

Si las preferencias del agente son completas, transitivas y continuas, podemos representarlas mediante una función de utilidad continua *u* que asigne un número *u*(A) a cada opción (A). En vez de decir que un agente

racional escoge la opción más factible, podemos decir entonces que el agente *maximiza la utilidad*. En esta frase, «utilidad» es un mero atajo para aludir a las preferencias con ciertas propiedades. Para verlo, podemos señalar que el único requisito para que una función u represente un orden de preferencias es que A se prefiera a B si y sólo si $u(A) > u(B)$. Si u siempre es positivo, $v = u$ también puede representar el mismo orden de preferencias, aunque v asigne números más grandes o (para $u < 1$) más pequeños que u. Los números absolutos carecen de significación; sólo la tiene su magnitud relativa u *ordinal*. Por ende, la idea de «maximización de la utilidad» no implica que el agente se dedique a conseguir lo máximo posible de algún «ingrediente» psíquico. Sí excluye, sin embargo, el tipo de jerarquía de valores encarnado en las preferencias lexicográficas. De hecho, éstas no pueden representarse mediante una función de utilidad.

Utilidad cardinal y actitudes ante el riesgo

A menudo, los agentes enfrentan opciones *de riesgo*, es decir elecciones que, con probabilidades conocidas, pueden tener más de un resultado posible. Intuitivamente, parecería que un agente racional ha de escoger la opción con la mayor *utilidad esperada*, una idea que incorpora la utilidad de cada resultado, así como su probabilidad de ocurrencia. Para cada opción, el agente ponderaría en principio la utilidad de cada consecuencia por su probabilidad y sumaría todas las utilidades ponderadas, y luego escogería la opción con la suma más grande.

La utilidad ordinal *no* nos permite, sin embargo, expresar en detalle esta idea. Supongamos que hay dos opciones, A y B. A puede producir el resultado O_1 o el resultado O_2 con probabilidades de ½ y ½, mientras que B puede producir los resultados O_3 u O_4, también con probabilidades de ½ y ½. Supongamos ahora una función de utilidad u que asigne valores 3, 4, 1 y 5 a O_1, O_2, O_3 y O_4, respectivamente. La «utilidad ordinal esperada» de A es 3,5, y la de B es 3. Si utilizamos, en cambio, la función $v = u$, los números son 12,5 y 13. Cada función representa las preferencias tan bien como la otra, no obstante lo cual ambas destacan diferentes opciones como «la mejor». Este enfoque es claramente inútil.

Es posible hacer algo mejor, pero con algunos costes conceptuales. El enfoque asociado a John von Neumann y Oskar Morgenstern muestra que

se pueden asignar a las opciones valores de utilidad que tengan una significación *cardinal* y no meramente ordinal. Un ejemplo de asignación de valores cardinales es la temperatura. El hecho de que la midamos en grados Celsius o Farenheit no afecta el valor de verdad del enunciado «la temperatura promedio en París es superior a la temperatura promedio en Nueva York». (Si las temperaturas se midieran ordinalmente, el enunciado carecería de sentido.) En contraste, el valor de verdad del enunciado «en París hace el doble de calor que en Nueva York» *sí* depende de la elección de la escala. Con todo, mientras que el valor de verdad de este enunciado específico sobre intensidades es sensible a la escala, otros no lo son. El valor de verdad del enunciado «la diferencia de temperatura entre Nueva York y París es más grande que entre París y Oslo», por ejemplo, no depende de la elección de la escala. De manera análoga, podemos construir medidas cardinales de utilidad que reflejen, entre otras cosas, como veremos, la intensidad de las preferencias y no sólo la clasificación ordinal de las opciones. Éstas nos permiten comparar la ganancia (o pérdida) de utilidad de ir de x a $(x + 1)$ con la de ir de $(x + 1)$ a $(x + 2)$, esto es, hablar de utilidad marginal creciente o decreciente, conceptos que carecen de significado para las mediciones de la utilidad ordinal.

No hace falta preocuparse por los detalles técnicos de la construcción, dado que la idea básica es simple y suficiente para nuestros fines presentes. Comenzamos por suponer que los agentes tienen preferencias no sólo en materia de opciones, sino también de *loterías* de opciones (incluidas las «loterías degeneradas» consistentes en obtener con certeza una opción básica). En cualquier conjunto dado de opciones básicas o «premios», una lotería especifica, para cada premio, la probabilidad de ganarlo; la suma de las probabilidades es 1. Se supone que los agentes tienen preferencias completas y transitivas con respecto a esas loterías. También se supone que las preferencias obedecen un «axioma de independencia»: la preferencia entre dos loterías p y q no se ve afectada si ambas se combinan de la misma manera con una tercera lotería r. El «efecto de certidumbre» mencionado en el Capítulo 7 y examinado con más detalle en el Capítulo 12 invalida ese axioma.

Para terminar, se supone que las preferencias exhiben una forma de continuidad, definida de la siguiente manera. Supongamos que las opciones básicas incluyen un mejor elemento A y un peor elemento B. Les asignamos, en forma arbitraria, números de utilidad 1 y 0. La continuidad

significa que para cualquier opción intermedia C hay una probabilidad p(C) tal que al agente le daría igual tener la certeza de conseguir C o participar en una lotería que le diera A con una probabilidad p(C) y B con una probabilidad 1 – p(C).[34] Definimos luego la *utilidad cardinal* u(C) como igual a p(C). Este número es arbitrario, claro está, porque lo son las utilidades terminales. Supongamos que asignamos los números de utilidad M y N a A y B, respectivamente (M > N). Definimos a continuación la utilidad de C como la utilidad esperada de la lotería:

$$pM + (1 - p)N = Mp + N - Np = (M - N)p + N.$$

La clase de las funciones de utilidad que aparecen de esta manera es mucho más pequeña que la clase de las funciones de utilidad ordinal.[35] Es fácil ver que si la opción X tiene mayor utilidad esperada que Y de acuerdo con una función, también la tendrá de acuerdo con cualquier otra. De tal modo, podemos afirmar sin caer en la ambigüedad que un agente racional maximiza la utilidad esperada.

Las funciones de utilidad cardinal tienen la importante propiedad de ser de *probabilidad lineal*. Introduzcamos la notación XpY, con referencia a una lotería que ofrece la probabilidad p de ganar X y 1 – p de ganar Y. Usando la escala valorativa 1-0, la utilidad u(X) es igual a la probabilidad q en la que el agente es indiferente entre X y la lotería AqB. De manera similar, la utilidad u(Y) es igual a la probabilidad r en la que el agente es indiferente entre Y y la lotería ArB. En consecuencia, XpY ofrece el equivalente de utilidad de una posibilidad p de obtener A con una probabilidad q, y una posibilidad 1 – p de obtener A con una probabilidad r. La utilidad de XpY, por lo tanto, es $pq + r(1 - p)$, que es p veces la utilidad de X más (1 – p) veces la utilidad de Y. Por ejemplo, la utilidad de una combinación probabilística de una posibilidad de 3/5 de obtener X y una posibilidad de 2/5 de obtener Y es $3/5q + 2/5r$.

Alguien podría hacer la siguiente objeción. Supongamos que un agricultor tiene que elegir entre dos cultivos: la variedad tradicional que tiene iguales probabilidades de producir una cosecha buena o mediocre, según el clima, y una variedad moderna que tiene iguales probabilidades de rendir una cosecha excelente o pobre. Supongamos asimismo que las utilidades cardinales son 3 y 2 para el viejo cultivo y 5 ó 1 para el nuevo.

Como la utilidad esperada del nuevo cultivo es más grande, el agricultor debería elegirlo. Pero, podría rezar la objeción, ¿no pasa esto por alto el hecho de que el agricultor siente tal vez aversión al riesgo y no está dispuesto a aceptar ninguna opción que pueda redundar en un nivel de utilidad tan bajo como 1? Sin embargo, esta objeción implica una doble contabilización, dado que la aversión al riesgo *ya* está incorporada a la construcción de las utilidades cardinales. Suponiendo que A, B y C tomen los valores de 100, 0 y 60, *u*(C) bien podría ser 0,75 para una persona adversa al riesgo, lo cual implica que le da lo mismo obtener 60 con certeza o intervenir en una lotería que le da una posibilidad del 25% de no ganar nada y una posibilidad del 75% de ganar 100. Un argumento similar es válido para la asignación de valores de utilidad cardinal a las cantidades materiales del cultivo.

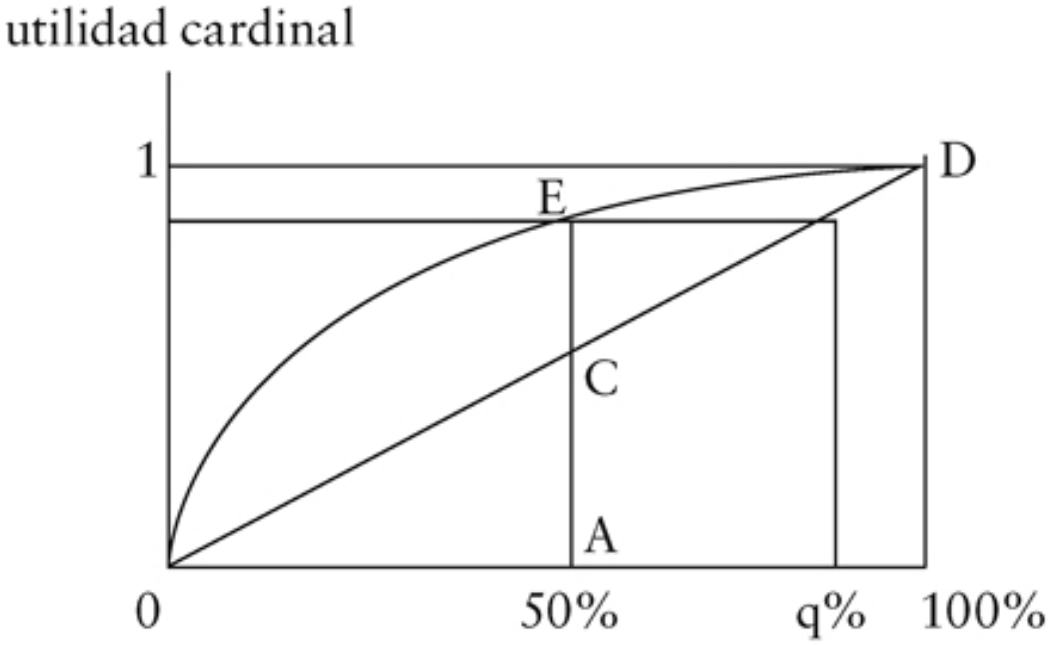


FIGURA 11.2

Consideremos otra ilustración, la adjudicación de la custodia de los hijos (véase la Figura 11.2). El eje horizontal puede entenderse de dos maneras, en cuanto implica o bien una división material de la custodia (porcentaje del tiempo pasado con el niño) o bien una división probabilística (la posibilidad de que un tribunal otorgue la custodia

exclusiva). La utilidad cardinal de la cantidad igual de tiempo compartido es AE, que es más grande que la utilidad AC de una posibilidad del 50% de custodia exclusiva. (Aquí apelamos al hecho de que la utilidad cardinal es de probabilidad lineal.) La razón es que, en esta situación, la mayoría de las personas se muestran remisas al riesgo. Están dispuestas a aceptar la custodia compartida porque un riesgo del 50% de no ver al hijo en absoluto es intolerable. El litigio sólo es preferible a la custodia compartida si un padre cree que su posibilidad de conseguir la custodia exclusiva es mayor que un porcentaje *q*. Si hay una considerable cantidad de litigios, no es porque a los padres les encante el riesgo, sino porque el pensamiento desiderativo les hace exagerar su posibilidad de recibir la custodia.

Aversión al riesgo y utilidad marginal decreciente

La exposición precedente, si bien exacta, puede ser engañosa. Parte de la literatura muestra una tendencia a difuminar la distinción entre aversión al riesgo y utilidad marginal decreciente. Para desarrollar este aspecto, necesito introducir un concepto que es intuitivamente valedero, aunque no se ha prestado (hasta ahora) a la medición. Se trata de la idea de la *utilidad intrínseca* de un bien, que refleja la intensidad de las preferencias del agente. La introspección nos dice de manera contundente que algunos bienes o experiencias brindan un inmenso disfrute, otros se limitan a ser satisfactorios, otros más son moderadamente fastidiosos y algunos, francamente espantosos. Representar la diferencia entre ellos simplemente en términos de preferencias ordinales –«prefiero el cielo al infierno, así como prefiero cuatro manzanas a tres»– equivale sin duda a emplear una noción muy empobrecida del bienestar o la utilidad. El hecho de que no haya una manera confiable de asignar números a niveles intrínsecos de satisfacción o insatisfacción no prueba que la idea no sea valedera, del mismo modo que nuestra incapacidad para cuantificar y comparar los niveles de satisfacción de diferentes individuos no demuestra que la idea de comparación interpersonal del bienestar carezca de sentido.

La idea de que muchos bienes tienen una utilidad marginal decreciente puede entenderse desde ese punto de vista. Para una persona pobre, los primeros dólares tienen gran utilidad, pero a continuación cada dólar adicional sucesivo pierde valor en términos subjetivos. Todos los fumadores saben que el primer cigarrillo de la mañana es el mejor, y que se disfruta

más cada cigarrillo si se regula el ritmo y no se los fuma con demasiada frecuencia. De hecho, el fumar un cigarrillo tiene dos efectos: produce gozo en el presente y reduce el disfrute de futuros cigarrillos.

Sin embargo, el segundo efecto no tiene que ser obligatoriamente negativo. Volvamos a considerar el caso de la custodia de los hijos. Para un padre, una tarde con el niño cada dos fines de semana puede resultar más frustrante que satisfactoria. Una tarde todos los fines de semana es más de dos veces más satisfactoria, porque los lazos emocionales más fuertes creados por la mayor frecuencia de los encuentros hacen que cada uno de éstos sea más satisfactorio. En el otro extremo del espectro temporal, la satisfacción extra de estar con el niño siete días a la semana y no seis supera la satisfacción extra de estar seis días y no cinco, porque la custodia exclusiva brinda los beneficios de la planificación irrestricta. En realidad, el hecho de estar con el niño tiene una utilidad marginal (intrínseca) creciente, según lo muestra la Figura 11.3.

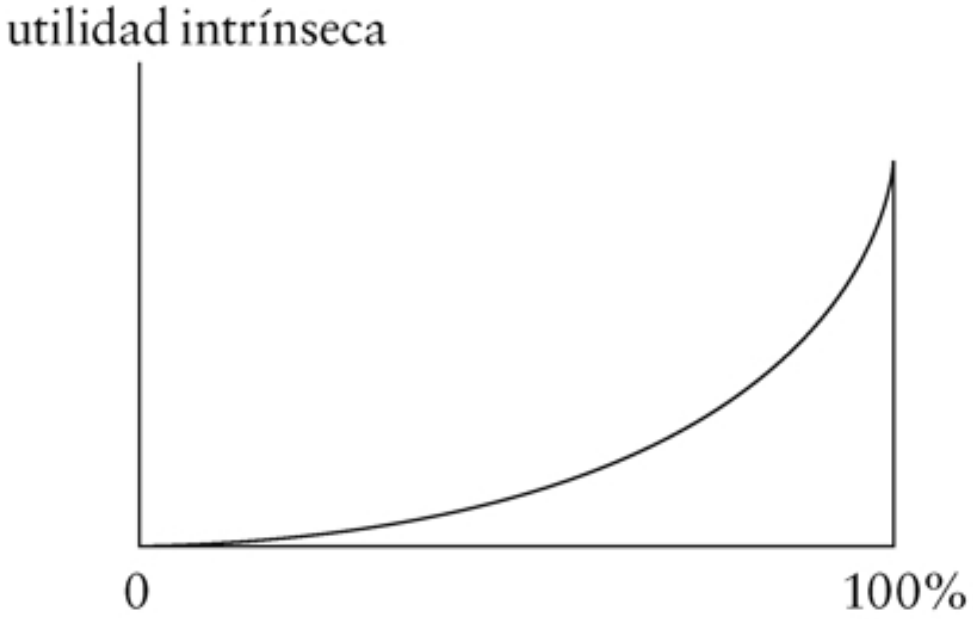


FIGURA 11.3

En este caso, la interpretación del eje horizontal es el porcentaje del tiempo pasado con el niño. Por las razones recién mencionadas, cada hora extra es más valiosa que la anterior. *Esta afirmación es perfectamente*

compatible con el análisis subyacente a la Figura 11.2. La utilidad marginal del tiempo pasado con el niño tal vez sea decreciente si se interpreta que la utilidad es cardinal, pero creciente si se la considera utilidad intrínseca. El hecho de que sólo la primera parte de este enunciado tenga una interpretación mensurable no implica que la segunda esté privada de sentido.

Si bien las funciones de utilidad cardinal siempre son generadas por dos factores psicológicos subyacentes, las actitudes ante el riesgo y la utilidad intrínseca, éstos no pueden medirse por separado. En rigor, no podemos decir si la curva OED de la Figura 11.2 se deriva de la neutralidad ante el riesgo combinada con la utilidad marginal intrínseca decreciente del tiempo pasado con el niño o de la aversión al riesgo combinada con la utilidad marginal intrínseca creciente de ese tiempo. En un caso dado, la intuición quizá nos diga cuál de las dos interpretaciones es más verosímil. Algunos padres pueden experimentar el tiempo pasado con el niño del modo como lo hacen muchos abuelos: es bueno en pequeñas dosis, pero pronto resulta agotador. A la vez, es posible que esos padres no se preocupen demasiado por el riesgo de no pasar absolutamente ningún tiempo con el niño (neutralidad del riesgo). Otros progenitores podrían diferir en ambos aspectos y generar así la misma función de utilidad cardinal. Pero repitámoslo, o volvamos a repetirlo: estos enunciados no pueden (hasta ahora) tener un carácter riguroso, aunque es obvio que tienen sentido.[36]

Creencias racionales

Aquí concluye el examen del primer componente de una elección racional: cómo escoger el mejor medio de realizar nuestros deseos, dadas nuestras creencias. Se destaca a todas luces que ésta es sólo una condición necesaria y no suficiente de la racionalidad. Si quiero matar a mi vecino y creo que la mejor manera de matar a alguien es hacer un títere que lo represente y clavarle alfileres, actúo racionalmente (en lo que concierne a este primer componente) si hago un títere que represente a mi vecino y le clavo alfileres. Sin embargo, salvo que existan circunstancias especiales, esa creencia difícilmente sea racional.[37]

Son racionales las creencias conformadas por el procesamiento de las pruebas disponibles mediante el uso de procedimientos que, a largo plazo y en promedio, son los que cuentan con mayores probabilidades de producir

creencias verdaderas. Supongamos que queremos formarnos una creencia acerca de la probabilidad de lluvias el 29 de noviembre, dentro de una semana. Tal vez no podamos hacer mucho más que examinar las estadísticas de precipitaciones en años anteriores y suponer que el futuro (esperado) será semejante al pasado. Pero a medida que el 29 de noviembre se acerca, las lluvias presentes quizá nos hagan modificar nuestras expectativas. Si en noviembre llueve con frecuencia, pero día tras día vemos el cielo sin nubes, es posible que infiramos la existencia de un sistema de alta presión que hace un tanto menos probable la lluvia el día 29.

Este proceso de revisión de creencias suele recibir el nombre de *aprendizaje bayesiano* (así llamado por Bayes, un ministro presbiteriano inglés del siglo XVIII). Supongamos que tenemos una distribución inicial («previa») de probabilidades subjetivas sobre distintos estados del mundo. En el ejemplo recién mencionado, la distribución previa se ha deducido de las frecuencias pasadas. En otros casos, podría tratarse de una mera corazonada. Sobre la base de la intuición, yo podría asignar, por ejemplo, una probabilidad del 60% a la competencia del primer ministro (PM) y un 40% a su incompetencia. Podemos observar luego las medidas que toma en el cargo y sus resultados, como el índice de crecimiento de la economía. Imaginemos que podemos hacer un cálculo acerca de la probabilidad de esas observaciones *dada la competencia* del PM. Con un PM competente tenemos un 80% de expectativas de buenos resultados; con un PM incompetente, sólo un 30%. Bayes mostró que podemos a continuación actualizar nuestras probabilidades iniciales en lo concerniente a la competencia del PM, *dadas las observaciones*.

Supongamos que hay sólo dos resultados posibles, bueno o malo, y que observamos uno bueno. Si escribimos *p(a)* la probabilidad de que *a* exista y *p(a \ b)* la *probabilidad condicional* de que *a* exista dado que *b* existe, hemos supuesto que *p*(PM es competente) = 60%, *p*(PM es incompetente) = 40%, *p*(buen resultado \ PM es competente) = 80% y *p*(buen resultado \ PM es incompetente) = 30%. Procuramos determinar *p*(PM es competente \ buen resultado). Utilizamos las letras *a* y *b* para indicar, respectivamente, competencia y buen resultado. En primer lugar escribimos, entonces, que

$$p(a \setminus b) = p(a \,\&\, b)/p(b) \qquad (*)$$

Dicho con palabras, la probabilidad condicional $p(a \setminus b)$ es igual a la probabilidad de que tanto *a* como *b* existan, dividida por la probabilidad de *b*. Esto se deduce de la idea más intuitiva de que $p(a \,\&\, b)$ es igual a $p(b)$ multiplicada por $p(a \setminus b)$. Si dividimos ambos lados de esta ecuación por $p(b)$, obtenemos la ecuación (*).

Si volvemos a utilizar la ecuación (*), pero con *a* y *b* invertidos, tenemos

$$p(b \setminus a) = p(a \,\&\, b)/p(a)$$

o, de manera equivalente,

$$p(a \,\&\, b) = p(b \setminus a) \cdot p(a)$$

Si reemplazamos la última expresión en (*), obtenemos

$$p(a \setminus b) = p(b \setminus a) \cdot p(a) \,/\, p(b) \qquad (**)$$

Ahora bien, hay dos maneras de que ocurra *b* (el buen resultado): con un PM competente o con un PM incompetente. Sobre la base del hecho de que la probabilidad de que uno de dos sucesos mutuamente excluyentes ocurra es la suma de las probabilidades de cada uno de ellos, podemos escribir

$$p(b) = p(b \,\&\, a) + p(b \,\&\, \text{no } a)$$

que, por el razonamiento del párrafo siguiente a (*), es equivalente a

$$= p(b \setminus a) \cdot p(a) + p(b \setminus \text{no } a) \cdot p(\text{no } a)$$

Si en (**) sustituimos $p(b)$ por esta expresión, obtenemos el *teorema de Bayes*:

$$p(a \setminus b) = p(b \setminus a) \cdot p(a)/[p(b \setminus a) \cdot p(a) + p(b \setminus \text{no } a) \cdot p(\text{no } a)]$$

La incorporación de las probabilidades numéricas en el lado derecho de esta ecuación nos dice que $p(a \setminus b) = 80\%$, es decir que la observación de un resultado exitoso eleva la probabilidad de que el PM sea competente del 60% al 80%. Una segunda y una tercera observaciones positivas la elevarían al 91% y luego al 97%.[38] Si otra persona calculó inicialmente que $p(a) = 0{,}3$ y no 0,6, tres observaciones positivas sucesivas elevarían su cálculo en primer lugar a 0,53, luego a 0,75 y por último a 0,89. Por lo tanto, tal vez no importe mucho si las corazonadas iniciales son poco confiables, dado que a medida que ingresa más y más información, las creencias actualizadas son cada vez más dignas de confianza. Con el paso del tiempo, las diferencias iniciales de opinión pueden quedar sumergidas bajo nuevas pruebas.[39] Para futuras referencias (Capítulo 23), también señalamos que cada nueva información tiene menos impacto que la anterior.

Inversión óptima en la recolección de información

El tercer componente de una acción racional es la inversión óptima de recursos, como tiempo o dinero, en la adquisición de más información. Como lo muestra la Figura 11.1, hay varios determinantes de ese óptimo. Primero, la cantidad de información que es racional recolectar depende de los deseos del agente.[40] Por ejemplo, un agente que no se preocupe mucho por las recompensas en el futuro distante no invertirá demasiado en la determinación del tiempo de vida esperado de un bien de consumo durable. De manera más evidente, es sensato acumular más información antes de tomar una decisión importante como la compra de una casa que al escoger entre dos botellas de vino del mismo precio. En este último caso, tal vez habría que decidir sencillamente mediante el lanzamiento de una moneda, si el coste esperado de determinar cuál es la mejor supera el beneficio esperado (basado en un conocimiento previo de la gama de calidad de vinos de igual precio) de beber el vino de superior calidad en vez del inferior.

Los deseos y las creencias previas determinan en conjunto los beneficios esperados de la nueva información. A veces es posible decir con gran precisión cuántas vidas más se salvarán si se realiza un análisis específico para el cáncer o, trasladado al nivel del agente, qué probabilidad hay de salvar su vida. El valor de la vida depende de cómo la negocie el agente en comparación con otros fines deseados. Según un cálculo, se necesitaba una prima de alrededor de doscientos dólares por año para inducir a hombres dedicados a ocupaciones riesgosas como la minería a aceptar una posibilidad en mil por año de sufrir una muerte accidental. En consecuencia, en la época en que se hizo ese cálculo, el valor de una vida era de unos doscientos mil dólares.[41] Los costes esperados de la nueva información, que están determinados por creencias previas, también pueden a veces estimarse con precisión. Para detectar un cáncer intestinal, es habitual realizar una serie de seis análisis poco costosos en las deposiciones de una persona. Los beneficios de los dos primeros análisis son significativos. Sin embargo, para cada uno de los últimos cuatro, los costes de detectar un caso adicional de cáncer (ni hablar de curarlo) se han calculado en 49.150, 469.534, 4.724.695 y 47.107.214 dólares, respectivamente.

La búsqueda óptima de información también puede depender de los resultados de la búsqueda misma (caso representado por el bucle de la Figura 11.1). Cuando se somete a prueba un nuevo producto médico, hay una decisión previa de suministrar la medicación a un grupo y no darlo a otro por cierto tiempo. Sin embargo, si no tarda en hacerse evidente que el producto tiene un éxito espectacular, sería poco ético abstenerse de suministrarlo al grupo de control. El mismo argumento es valido para un solo agente racional. Supongamos que voy al bosque a recoger bayas. Sé que éstas tienden a crecer en racimos, de modo que estoy preparado para dedicar algún tiempo a buscar antes de empezar a recogerlas. Si tengo suerte y encuentro al principio un lugar donde abundan, sería tonto seguir buscando.

Podemos ver la recolección de información como una *acción en la sombra* que acompaña la acción primaria. Antes de decidir qué hacer, tenemos que decidir cuánta información vamos a recoger. A veces, *la acción en la sombra y la acción primaria pueden coincidir*, al menos en parte. Supongamos que los líderes de un país estudian la posibilidad de ir a la guerra contra otro país. La invasión alemana de Francia, en 1940, puede

servirnos de ejemplo. Para tomar la decisión final de atacar, la información era crucial. Las autoridades necesitaban conocer las capacidades objetivas del enemigo previsto, así como «la organización, las costumbres y los hábitos del ejército del enemigo» (frase tomada de los *Deberes del Estado Mayor*). Gran parte de esa información podía reunirse por medios convencionales, incluido el espionaje. Sin embargo, para determinar la *moral* del enemigo (su espíritu de combate), no había otra alternativa que trabarse concretamente en lucha con él.

Indeterminación

Estos últimos ejemplos, la recolección de bayas y la planificación bélica, también nos ayudarán a ver las *limitaciones* de la teoría de la elección racional o, mejor, una de sus dos limitaciones. Como herramienta explicativa, la teoría puede fallar en uno de dos aspectos. Por un lado, puede no producir predicciones únicas sobre lo que hará la gente en una situación dada. Por otro, la gente tal vez no logre mantenerse fiel a sus predicciones, sean éstas únicas o no. La segunda falla, la *irracionalidad*, es el tema del próximo capítulo. La primera, la *indeterminación*, es el tópico de las siguientes observaciones.

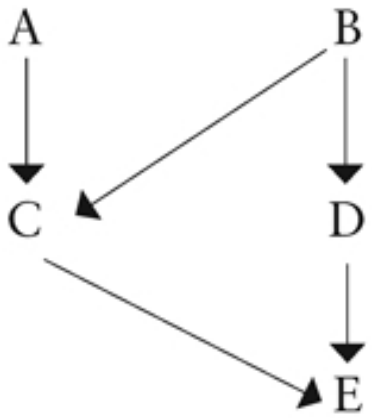


FIGURA 11.4

Un agente puede ser incapaz de identificar el mejor elemento en el conjunto conveniente, por una de dos razones. Un consumidor tal vez sea

indiferente entre dos opciones que son igualmente óptimas. En casos triviales, esto sucede cuando las opciones son imposibles de distinguir, por ejemplo cuando un consumidor debe elegir entre dos latas idénticas de sopa en el supermercado. En casos no triviales, dos opciones podrían diferir en varias dimensiones, de modo que las diferencias se compensan exactamente unas a otras. El caso no trivial es raro y quizás inexistente. Si se me propone elegir entre dos automóviles que difieren en precio, confort, apariencia, velocidad, etc., tal vez no prefiera ninguno en desmedro del otro sin que, para el caso, ambos me den igual. Si me dieran igual, un descuento de cinco centavos en uno de los autos debería inducir una preferencia por esa opción. La intuición sugiere que es improbable que suceda así.

De hecho, las preferencias del consumidor pueden ser *incompletas*. Supongamos que he examinado cinco modelos de autos, A, B, C, D y E, y los he clasificado como se muestra en la Figura 11.4 (las flechas representan la relación de preferencia). Mi incapacidad para comparar C y D no importa, dado que, de todos modos, no voy a comprar ninguno de los dos. En contraste, la incapacidad para comparar A y B me pone en un aprieto. Es cierto, podría tratar de recoger más información, pero ¿cómo sé que el esfuerzo merece la pena? Dentro de un instante vuelvo a este punto. Querría antes señalar otro origen de las preferencias incompletas, probablemente más importante. De ordinario, las preferencias de resultados inducen las preferencias de opciones. Prefiero una opción porque prefiero su resultado, esto es, su utilidad esperada, en comparación con la de las demás. Sin embargo, si la situación es de incertidumbre o ignorancia y no de riesgo (Capítulo 7), tal vez no sea capaz de comparar los resultados.[42] Según las inmortales palabras del doctor Johnson:

> La vida no es larga, y no es bueno que una parte excesiva de ella transcurra en ociosas deliberaciones sobre cómo pasarla; la deliberación, en aquellos que la inician con prudencia y la prosiguen con sutileza, debe, tras prolongado dispendio de reflexiones, concluir en el azar. El preferir un modo futuro de vida a otro, sobre la base de atinadas razones, exige facultades que nuestro Creador no ha tenido a bien otorgarnos.[43]

Otra indeterminación aparece en la dificultad de determinar la inversión óptima en la recolección de información. Cuando salgo a recoger bayas en un territorio desconocido, ¿cuánto tiempo debo dedicar a buscar un racimo denso y cuándo debo comenzar a recoger? A menos que encuentre de inmediato una parcela abundante, es sensato dedicar algún tiempo a hacer una recorrida exploratoria. A la vez, no quiero seguir buscando hasta la

caída de la noche, porque en ese caso, sin lugar a dudas, volveré a casa con la canasta vacía. Entre los límites superior e inferior del tiempo que uno debe dedicar a buscar, puede haber un gran intervalo de indeterminación. Un problema diferente se plantea en el caso de la planificación para la guerra. Si la decisión primaria y la decisión en la sombra coinciden, el planificador está condenado a permanecer en un estado de incertidumbre (al menos parcial). La teoría de la elección racional no puede orientarnos bien en esas situaciones. Es útil en situaciones sumamente estructuradas sobre las cuales se conoce mucho, como los análisis para detectar el cáncer, pero no tanto en entornos desconocidos.

Así como el gasto de menos tiempo que el indicado por el límite inferior y de más tiempo que el marcado por el límite superior sería irracional, ninguna elección que el agente haga *dentro* de ese intervalo podrá caracterizarse de ese modo. En consecuencia, se podría, tal vez, pensar en abandonar la idea de racionalidad en beneficio de la idea de *no irracionalidad*. Esta versión revisada de la teoría nos permitiría comprender una gama más amplia de comportamientos, pero tener menos capacidad predictiva. La mayoría de quienes la practican serían renuentes, me parece, a revisarla de esa manera. Lo que les atrae ante todo de ella es precisamente su promesa de generar predicciones *únicas*. Lo hace en virtud del hecho matemático elemental de que cualquier función de utilidad «obediente» definida con respecto a un conjunto de oportunidades «obedientes» alcanza un valor máximo para un único miembro de ese conjunto. La interacción entre el conjunto de oportunidades y las curvas de indiferencia de la Figura 9.1 es un buen ejemplo de la atrapante simplicidad de la idea de «hacer lo mejor posible».

La fuente más importante de indeterminación de las creencias se da en la interacción estratégica, cuando cada agente tiene que formarse creencias acerca de lo que los otros probablemente han de hacer sobre la base de las *suyas propias*, sabiendo que apelan a un razonamiento similar al suyo. En algunos casos, examinados con mayor detalle en el Capítulo 19, la estructura de recompensas no permite a los agentes converger en un conjunto de creencias compartidas.

La racionalidad es subjetiva de uno a otro extremo

Querría concluir este examen de la teoría de la elección racional volviendo a destacar su *naturaleza radicalmente subjetiva*. Uno podría, desde luego, tomar la palabra «racional» en un sentido objetivo, y dar a entender que un agente racional toma decisiones que *mejoran su vida*, según es posible juzgarlo con criterios objetivos como la salud, la longevidad o el ingreso. Utilizada de este modo, sin embargo, la idea no tendría ninguna capacidad explicativa. Como he señalado, las *consecuencias* de una decisión no pueden explicarla. Sólo los estados mentales que la preceden nos permiten *explicar* las acciones como óptimas desde el punto de vista del agente en vez de *caracterizarlas* como útiles o beneficiosas desde el punto de vista de un observador externo (o del agente en un momento posterior).

Supongamos que padezco de una grave incapacidad para posponer la gratificación, es decir que no puedo tomar en cuenta las consecuencias futuras del comportamiento presente. Y supongamos además que los científicos dan con una píldora del descuento, que incrementa el peso de las recompensas futuras en las decisiones presentes. Si tomo la píldora, mi vida mejorará. Mis padres estarán contentos si la tomo. En retrospectiva, estaré agradecido por haberla tomado. Pero de tener la opción de tomarla o no, si soy racional la rechazaré. Cualquier conducta que la píldora pueda inducir ya está a mi alcance. Podría dejar de fumar, comenzar a hacer ejercicios o empezar a ahorrar ya mismo, pero no lo hago. Como no quiero hacerlo, no me gustaría tomar una píldora que me lo impusiera. De manera similar, una persona egoísta rechazaría una «píldora del altruismo» y, en forma aún más ineludible, una persona altruista se negaría a tomar una «píldora del egoísmo». Si amo a mi familia y estoy dispuesto a sacrificar parte de mi bienestar hedonista por su bien, rechazaría una píldora cuyo efecto de dos pasos consistiera en reducir su bienestar, así como desestimaría cualquier opción (por ejemplo, pagarme un plato costoso) que produjera el mismo efecto en un paso.

Para aguzar el argumento, supongamos que una persona consume x hoy e y mañana, y que su tasa de descuento uniperiódica es de 0,5 (la persona es indiferente entre una unidad de utilidad mañana y media unidad hoy). Y supongamos, para simplificar, que $u(x) = x$, y $u(y) = y$. El valor descontado presente de su flujo de consumo es $x + 0{,}5y$. Imaginemos que la persona se entera de que mañana va a sufrir un dolor que reducirá la utilidad de su consumo por un factor de 0,5. El valor descontado presente del consumo es

ahora $x + 0{,}25y$. Si un agente racional recibe el ofrecimiento de una aspirina sin coste que eliminará el dolor, sin duda la aceptará y de ese modo restablecerá el valor presente original. Si tomara una píldora que indujera una tasa de descuento de 1 (pero no aceptara la aspirina), el resultado sería el mismo, en el sentido de que, en uno y otro caso, sería indiferente entre el flujo biperiódico y una utilidad uniperiódica de $x + 0{,}5y$. Como el agente tomaría la aspirina y puesto que ésta tiene el mismo efecto que la píldora del descuento, ¿por qué no tomaría esta última? La razón es que la elección de esa píldora está restringida por la necesidad de que el consumo inducido por ella sea superior al consumo sin ella, *según se juzga de conformidad con las preferencias previas a la aparición de la píldora*. No hay una limitación similar en la elección de la aspirina, porque no hay diferencia entre las preferencias anteriores y posteriores a ella. Aun sin la aspirina, prefiero poder liberarme del dolor mañana. Cuando ese estado se convierte en parte de mi repertorio, elijo darle lugar. En contraste, el flujo de utilidad inducido por la píldora del descuento ya está en mi repertorio, pero decido no darle lugar.[44]

En otras palabras, las elecciones deben verse a través de los ojos del agente. Una persona miope que pierde sus gafas puede estar impedida de encontrarlas justamente a causa de su miopía. Está «atrapada».[45] Demanera similar, un agente racional quizá se encuentre en una «trampa de creencias» que lo deja pegado a una falsa creencia, si cree que los costes de someterla a prueba son demasiado altos. Así, las mujeres que practican la mutilación genital pueden estar atrapadas en esa situación. Los bambaras de Mali creen que el clítoris matará al hombre si tiene contacto con el pene durante el coito. En Nigeria, algunos grupos creen que si la cabeza del bebé toca el clítoris durante el parto, la criatura morirá. En Polonia tuvo amplia circulación la creencia de que las bebidas alcohólicas provocarían la muerte de las personas que tienen implantado disulfiram (Antabuse) debajo de la piel. En realidad, el disulfiram implantado es farmacológicamente inerte. No obstante, la falsa creencia podría disuadir a la gente de hacer la prueba.

La *racionalidad* de las creencias es algo completamente diferente de su *verdad*. En tanto que la verdad es un rasgo de la relación entre la creencia y el mundo, la racionalidad es un rasgo de la relación entre la creencia y las pruebas en poder del agente. Aunque la racionalidad exija al agente invertir en nueva información, la inversión siempre está restringida por sus costes y beneficios esperados (es decir, *creídos*). Si se cree que la recolección de

mayor información tiene elevados costes de *oportunidad*, como cuando se está frente a un peligro inminente, quizá sea racional no hacer la inversión. Si se cree que tiene altos costes *directos*, como cuando se somete a prueba la creencia sobre los efectos fatales de la bebida en quienes usan un implante de disulfiram, sólo una persona irracional se arriesgaría a invertir. En términos más generales, muchas creencias deben tomarse por lo que valen en forma indirecta, pues si las pusiéramos todas a prueba, jamás seguiríamos adelante con nuestra vida.

Cualquier explicación del comportamiento basada en la elección es subjetiva. Sin embargo, no todas las explicaciones subjetivas suponen la transparencia de los agentes para sí mismos y la búsqueda inflexible de la optimalidad, marcas distintivas de las explicaciones fundadas en la elección racional. En el capítulo siguiente examinaré una serie de explicaciones que se apartan de la teoría de la elección racional en uno de esos aspectos o en ambos.

* * *

NOTA BIBLIOGRÁFICA

Analizo la relación entre razón (en el sentido expuesto en el Capítulo 4) y racionalidad en mi lección inaugural en el Collège de France, *Raison et raisons*, París, Fayard, 2006. Quien busque más sobre Weber y la racionalidad puede consultar mi «Rationality, economy, and society», en Stephen Turner (comp.), *The Cambridge Companion to Weber*, Cambridge y Nueva York, Cambridge University Press, 2000, págs. 21-41. Se hallará una exposición clásica de la teoría de la utilidad en R. Duncan Luce y Howard Raiffa, *Games and Decisions: Introduction and Critical Survey*, Nueva York, Wiley, 1957. La obra original de John von Neumann y Oskar Morgenstern, *The Theory of Games and Economic Behavior*, 2ª edición, Princeton, Princeton University Press, 1947, aún merece consultarse. Reid Hastie y Robyn Dawes, *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*, Thousand Oaks (California), Sage, 2001, presentan una sobresaliente exposición de la teoría de la elección racional (y sus problemas). Analizo con mayor detalle el ejemplo de la custodia de los hijos en el Capítulo 3 de *Solomonic Judgments: Studies in the Limitations of Rationality*, Cambridge y Nueva York, Cambridge University Press, 1989 [*Juicios salomónicos: las*

limitaciones de la racionalidad como principio de decisión, Barcelona, Gedisa, 1991]. Robert L. Winkler, *An Introduction to Bayesian Inference and Decision*, Gainesville (Florida), Probabilistic Publishing, 2003, es una excelente presentación básica de la teoría bayesiana. Conversaciones con Gary Becker y Peter Diamond han influido en mi argumento de que una persona racional no tomaría la píldora del descuento; véase también Ole-Jørgen Skog, «Theorizing about patience formation: the necessity of conceptual distinctions», *Economics and Philosophy*, 17, 2001, págs. 207-219. Tomo la idea de la trampa de creencias de Gerry Mackie, «Ending footbinding and infibulation: a convention account», *American Sociological Review*, 61(6), 1996, págs. 999-1017. Se encontrará un útil estudio de la importancia de la inteligencia en la preparación de la guerra en Ernest R. May, *Strange Victory: Hitler's Conquest of France*, Nueva York, Hill and Wang, 2000. Debo la información sobre el uso de implantes de disulfiram en Polonia a Wiktor Osiatynski, *Alcoholism: Sin or Disease?*, Varsovia, Stefan Batory Foundation, 1997, y los datos acerca de su ineficacia a Jon Johnsen y Jørg Mørland, «Depot preparations of disulfiram: experimental and clinical results», *Acta Psychiatrica Scandinavica*, 86(369), 1992, págs. 27-30.

Notas al pie

[29] De hecho, el segundo sesgo de los fumadores no compensa del todo el primero.

[30] En las ratas, la demora entre la respuesta irreflexiva y la respuesta reflexiva es de alrededor de diez milisegundos.

[31] Ignoro la condición técnica de que el conjunto de opciones sea compacto y cerrado. La compacidad queda invalidada si para cualquier opción factible hay otra que es mejor, y el cierre, si el límite mismo de una secuencia de opciones factibles no es factible.

[32] En la nota 1 del Capítulo 18 se encontrará otra manera de «mejorar hasta morir».

[33] En consecuencia, la noción intuitiva de incomparabilidad puede expresarse de dos maneras distintas: como preferencias incompletas o como preferencias discontinuas.

[34] La *identificación* de esta probabilidad plantea el problema del anclaje mencionado en la introducción a la segunda parte.

[35] Cualquiera de esas dos funciones están, de hecho, relacionadas entre sí como las escalas de temperatura de Celsius y Farenheit, que asignan diferentes valores (correspondientes a M y N en el texto) a las temperaturas en las que el agua hierve y se congela.

[36] Por ende, la analogía con las escalas de temperatura sólo es parcialmente válida. Estas escalas *sólo* miden la intensidad de la temperatura. Las funciones de utilidad cardinal miden el resultado conjunto de la intensidad de la preferencia y las actitudes ante el riesgo.

[37] La creencia en la brujería puede autocumplirse si la persona maldecida cree en la eficacia de la maldición y simplemente pierde la voluntad de vivir. En ese caso, la eficacia observada de la maldición podría hacer que la creencia en la brujería fuera racional, aun cuando (como sucede con la teoría de la acción a distancia) el agente no pudiera especificar su mecanismo de funcionamiento. También podría hacer punible la brujería sobre la base de sus consecuencias reales y no, como lo sugieren Donne y Hobbes (véase la introducción a la segunda parte), exclusivamente sobre la base del dolo.

[38] En este razonamiento nos basamos en dos clases de creencias previas, una «creencia de primer plano» concerniente al tipo del PM y una «creencia de segundo plano» sobre la probabilidad de que un PM de un tipo dado produzca un buen resultado. Mientras que la creencia de segundo plano se mantiene constante en todo momento, la creencia de primer plano se revisa conforme recibimos nueva información. También podríamos trasladar el primer plano al segundo y a la inversa, si lo que nos interesa es determinar si los políticos competentes son, en los hechos, capaces de dar forma a los resultados económicos o éstos son principalmente afectados por factores aleatorios o exógenos. Podríamos entonces comenzar con una creencia de primer plano acerca de la probabilidad de que los buenos políticos produzcan buenos resultados y una creencia de segundo plano sobre la competencia de un político determinado. Si la observación muestra que la economía funciona mal bajo la conducción de un político a quien se asigna una alta probabilidad de ser de tipo competente, podríamos atribuir una mayor probabilidad a la hipótesis de las influencias aleatorias o exógenas.

[39] Para que haya convergencia, las nuevas informaciones sucesivas deben ser estadísticamente independientes entre sí. En el ejemplo de manual de la formación de creencias bayesianas, una persona extrae bolas de una urna a cuyo respecto se sabe que es igualmente probable que contenga un 80% de negras y un 20% de blancas o un 20% de negras y un 80% de blancas, a fin de determinar la probabilidad de que exista uno u otro estado. Como las extracciones se realizan al azar y después de cada una de ellas las bolas se devuelven a la urna, el resultado de cada extracción es independiente de los anteriores. En situaciones políticas como la descrita en el texto, puede ser mucho más difícil verificar la independencia. Además, la convergencia presupone que la situación subyacente sigue siendo la misma o al menos no cambia con demasiada celeridad. Si el entorno se modifica

rápidamente, el proceso de actualización de las creencias se parece al de apuntar a un blanco móvil (Capítulo 6).

[40] El pensamiento desiderativo, en el que «el deseo es el padre del pensamiento», es claramente irracional. En contraste, no hay nada de irracional en el proceso mostrado en la Figura 11.1, en el cual los deseos son, por así decirlo, los abuelos de las creencias.

[41] Hay muchos escollos a la hora de hacer esos cálculos, pero el sentido general es imposible de negar: todos atribuimos un valor finito a nuestra vida. Si no lo hiciéramos, no nos dedicaríamos a todas las actividades riesgosas agradables o rentables a las que nos dedicamos.

[42] Al tomar decisiones en condiciones de incertidumbre, quizá pueda comparar opciones si el peor resultado de una de ellas es mejor que el mejor de otra. En la toma de decisiones en condiciones de ignorancia, ni siquiera tenemos acceso a esa modesta posibilidad de comparar.

[43] Yo podría, no obstante, «concluir en el azar» y luego inventar las «atinadas razones», por ejemplo, para dar mayor peso a los atributos en los cuales la opción elegida es claramente superior. Este proceder puede tener consecuencias indeseables. Supongamos que debo elegir entre ir a la Facultad de Derecho o a la Facultad de Ciencias Forestales. Al ser incapaz de tomar una decisión fundada en la razón, voy a la Facultad de Derecho más o menos por azar y justifico la decisión de manera retrospectiva, atribuyendo más peso a la dimensión de ingresos de las dos carreras. Con estas preferencias recién inducidas, podría seguir adelante y tomar otras decisiones que difirieran de las que habría tomado sobre la base de mis preferencias anteriores a la elección.

[44] Con el descuento hiperbólico, un agente podría aceptar una píldora del descuento. Mediante el ejemplo numérico del Capítulo 6, supongamos que el efecto de la píldora consiste en reducir el valor de k de 1 a 0,3. En el momento en que la recompensa más pequeña está disponible, su valor presente es simplemente 10 (no hay descuento). El valor presente de la recompensa más grande en ese momento es $30/(1 + 0{,}3 \cdot 5) = 12$. Por lo tanto, el precompromiso bajo la forma de la ingesta de la píldora permitirá al agente actuar de conformidad con su juicio calmo y reflexivo e impedirá así la debilidad de la voluntad (en el sentido general). Este enunciado sigue siendo cierto aun cuando tenga que comprar la píldora, siempre que el coste de ésta (en términos de utilidad) sea inferior a 2. También sigue siendo cierto si el precompromiso tiene el efecto de reducir el valor de la recompensa postergada (tal vez la píldora del descuento tenga el efecto secundario de reducir la capacidad de disfrute), siempre que la pérdida sea inferior a 5. Estos hechos podrían ser relevantes si sustituimos la píldora del descuento por la psicoterapia.

[45] Si se le ofrecieran sus gafas, se las pondría. He sostenido que si se le ofreciera la píldora del descuento, no la tomaría. La diferencia es que ya puede hacer sin ella todo lo que podría hacer si la tomara, mientras que hay muchas cosas que no puede hacer sin las gafas que podría hacer si las tuviera puestas.

Capítulo 12
Racionalidad y comportamiento

La ignorancia de los costes de la toma de decisiones

La idea de racionalidad tiene un fuerte atractivo normativo. *Queremos* tener razones (deseos y creencias a la luz de los cuales la acción se muestre racional) para lo que hacemos. De hecho, nuestro deseo de actuar por una razón, nuestra deferencia a la racionalidad, puede tener el vigor suficiente para inducir un comportamiento irracional.[46] Podemos definir la *hiperracionalidad* como la propensión a buscar la decisión óptima en abstracto, es decir, la que sería óptima si ignoráramos los costes del proceso mismo de toma de decisiones. Estos costes son de tres clases: 1) el coste de los medios de decidir; 2) el coste de los efectos secundarios de decidir, y 3) los costes de oportunidad de decidir, esto es, el valor de las otras cosas que podríamos haber hecho en vez de recorrer el proceso decisional. Querría ilustrarlos brevemente.

La hiperracionalidad debido a la ignorancia de 1) podría surgir en las *compras por comparación*, cuando los ahorros (esperados) provenientes del descubrimiento del precio más bajo son menos que el dinero gastado en los viajes de tienda en tienda. En el sur de Francia, los turistas cruzan la frontera para comprar cigarrillos baratos en España, como si la gasolina fuera gratis. La ignorancia de 2) podría inducir una hiperracionalidad en los casos litigiosos de custodia de los hijos. El tribunal quizá trate de promover el interés del niño determinando cuál de los progenitores es más apto para recibir la custodia.[47] Una vez zanjada esa cuestión, el juez tiene una buena razón para otorgar la custodia a ese padre. Sin embargo, en el proceso jurídico

psicológico para verificar la aptitud relativa se puede infligir un daño incalculable al niño. Visto el objetivo que debe alcanzarse, un procedimiento más racional podría consistir en lanzar una moneda o atenerse a la presunción tradicional de la custodia materna.

La ignorancia de los costes de oportunidad se ilustra con una observación del doctor Johnson en una conversación con Boswell, sobre los temas que era preciso enseñar en primer lugar a los niños:

> Señor, importa tan poco qué tema les enseñéis el primero como qué pierna ponéis la primera en vuestros pantalones. Señor, por mucho que perseveréis en discutir cuál es más adecuado poner antes, entretanto vuestras posaderas estarán desnudas. Señor, mientras consideráis cuál de dos cosas deberíais enseñar la primera a vuestro hijo, otro niño aprende ambas.[48]

Otra vez, quizá sea más racional lanzar una moneda. Considérese asimismo el caso del médico que llega a la escena de un accidente y tienen que decidir qué medidas tomar. Aunque es obvio que necesita examinar al paciente, su comportamiento es contraproducente si le dedica tanto tiempo que éste muere en sus brazos («La operación ha sido un éxito, pero el paciente ha muerto»). Otros pueden haber tenido la experiencia, mientras recogían bayas, de buscar tanto tiempo el mejor lugar que lo hallaron cuando ya caía la noche. Aunque los ahorros obtenidos con las compras por comparación superen los costes de transporte, el comportamiento podría, con todo, seguir siendo irracional debido al valor del tiempo perdido.

Algunos principios canónicos de racionalidad

Al hablar de «enigma» entenderé aquí un comportamiento observado que parece recalcitrante a la explicación basada en la elección racional. Si bien algunos enigmas pueden, luego de una inspección más detenida, perder su carácter desconcertante, muchos lo mantienen. Los experimentos y las conductas en la vida real muestran numerosos ejemplos de patrones de conducta que violan los cánones de la racionalidad. En la siguiente lista selectiva de esos principios canónicos, comienzo con los más fundamentales y de allí voy a los más específicos. Me limito a las elecciones individuales; las anomalías en las elecciones interactivas se examinan en la quinta parte. La lista puede ser complementada con algunas de las anomalías cognitivas que discutí en el Capítulo 7.

1. En una elección entre actuar y no hacer nada, un agente racional no actuará si los costes de utilida esperados de la actuación superan los beneficios esperados de la utilidad.

2. En la elección entre males, un agente racional elegirá el mal menor.
3. Un agente racional asigna el mismo peso a los costes de oportunidad y los costes directos.
4. Un agente racional nunca preferirá tener un subconjunto de un conjunto de opciones a tener el conjunto completo.
5. Si un agente racional prefiere X a un vaso descrito como medio lleno, también debe preferir X a uno descrito como medio vacío.
6. En un juego de puro azar, un jugador racional, al hacer sus apuestas, no prestará atención a los resultados de las jugadas previas.
7. Cuando decide si persistir en un proyecto o arrumbarlo, un inversor racional sólo presta atención a valor presente de los flujos de utilidad futura de estas dos opciones.
8. Si en un tiempo 1 un agente racional planea llevar a cabo la acción X en el tiempo 2, hará X en es tiempo 2 a menos que, en el ínterin, sus deseos o creencias hayan cambiado.
9. En una elección riesgosa, un agente racional escogerá medios acordes con el resultado esperado, n sólo de conformidad con la mejor (o la peor) perspectiva.
10. En un mercado de agentes racionales, la tasa de retorno de todos los activos debería ser (aproximadamente) la misma.[49]
11. Si un agente racional escoge A del conjunto (A, B, C), también escogerá A del conjunto (A, B).
12. Un agente racional no actuará sobre un efecto para suprimir la causa (para curar una neumonía, tomará antibióticos y no aspirina).
13. Si un agente racional prefiere recibir la recompensa X con toda certeza en vez de recibir la recompensa Y con una probabilidad *q*, también preferirá recibir X con una probabilidad *p* a recibi Y con una probabilidad *pq* (axioma de independencia de la teoría de la utilidad cardinal).
14. Si un agente racional hace X cuando sabe que existe la circunstancia C (o pretende hacer X cuand se espera que C exista) y hace X cuando la circunstancia C no existe (o pretende hacer X cuando n se espera que C exista), debe hacer o pretender hacer X aun cuando desconozca las circunstancias
15. Un agente racional nunca hará un ofrecimiento si su aceptación revela información que genera un valor esperado negativo en el trato.
16. Si una ofensa suscita un deseo de venganza, la persona ofendida, de ser racional, se tomará su tiempo hasta que pueda devolver el golpe con una probabilidad máxima de éxito o un riesgo mínimo para sí misma.[50]
17. Si se lo desafía a librar un duelo de espadas, un agente racional tomará clases de esgrima en caso tener que aceptar el reto.
18. Antes de pedir la mano de otra persona, un agente racional recogerá información sobre sus propensiones de comportamiento y emocionales.

Transgresiones del canon

Según resultan las cosas, estos principios normativamente ineludibles son objeto de una transgresión constante. Siguen los ejemplos (cuyos números corresponden a los de los principios que transgreden).[51]

1. *La paradoja de la votación*. Como ninguna elección nacional se ha ganado nunca por un solo vot el sufragio de un individuo no marca diferencia alguna en el resultado y puede entrañar una considerable molestia para el votante. Pese a ello, gran cantidad de gente acude a las urnas.[52]
2. *Un mayor dolor preferido a un menor dolor*. Los sujetos de un experimento fueron expuestos a d

secuencias de ruido muy desagradable. Ambas secuencias implicaban la exposición a treinta y cin segundos de ruido de alto nivel. En la primera, seguían a esa exposición quince segundos de ruidc gradualmente decreciente (aunque todavía desagradable). Cuando se les preguntó a qué secuencia preferirían volver a exponerse, los sujetos escogieron la que era sin lugar a ninguna duda menos agradable.

3. *La paradoja del corte del césped*. En una pequeña comunidad suburbana, el señor H corta su césped. El hijo de su vecino lo haría por doce dólares. H no cortaría la parcela de césped de su vecino, que es del mismo tamaño, por veinte dólares.
4. *El enigma del club de Navidad*. En este sistema, los clientes depositan una suma mensual sin inter o con un interés muy bajo, que sólo pueden retirar en Navidad. La opción de recibir un interés normal y hacer retiros sin coste también está a su alcance.
5. *La paradoja de la tarjeta de crédito*. Cuando se introdujeron las tarjetas de crédito en el mercado, quienes las promovían preferían que cualquier diferencia entre el precio en efectivo y el precio co tarjeta se considerara como un descuento por pagar en efectivo y no como un recargo por pagar c tarjeta. Aunque las dos opciones son lógicamente equivalentes, los clientes se inclinaban más a utilizar las tarjetas si la diferencia se planteaba como un descuento por pago en efectivo.
6. *Las falacias de los dos jugadores*. Si el rojo ha salido cinco veces seguidas, alrededor de la mitad los jugadores cree que hay una probabilidad de más del 50% de que la próxima vez salga el negro La otra mitad cree que la probabilidad de que salga el negro es de menos del 50%.
7. *La falacia del coste hundido*. «La cancelación de un proyecto en el cual se han invertido mil cien millones de dólares representa un mal manejo exorbitante del dinero de los contribuyentes» (senador Denton, 4 de noviembre de 1981). En ocasiones, esta falacia también se denomina «falac del Concorde», por el costoso proyecto anglofrancés del avión de ese nombre, o «falacia de Vietnam», por la renuencia de los Estados Unidos a retirarse de ese país. Si compramos entradas para un espectáculo y una nevada abundante hace muy oneroso llegar hasta el lugar, tal vez decidamos ir pese a todo, aun cuando habríamos rechazado las entradas si nos las hubieran ofreci gratis.
8. *El enigma del dentista*. El 1º de marzo concierto una cita con la dentista para el 1º de abril. El 30 marzo la llamo para decirle que, debido a un funeral (ficticio) en la familia, no puedo ir. Salvo por el mero transcurso del tiempo, entretanto no se ha producido ningún cambio. En particular, el dol de muelas es el mismo.
9. *El mejor y el peor escenarios*. Los pacientes de cáncer en las últimas etapas de la enfermedad a menudo sobrestiman sus posibilidades de supervivencia. En vez de una terapia paliativa para mitigar el dolor, prefieren una agresiva y dolorosa quimioterapia con escasos beneficios. Cuando les pregunta cuánto pagarían para disminuir las posibilidades de que ocurra un desastre de baja probabilidad, las personas se muestran dispuestas a pagar la misma suma para reducirlas tanto a u en un millón como a una en diez millones.
10. *El enigma de la prima de los activos bursátiles*. Históricamente, el rendimiento de las acciones es superior al rendimiento de los bonos. Una persona que hubiese invertido un dólar en acciones el 1 de enero de 1928, habría tenido una cartera valuada en mil ochocientos dólares el 1º de enero de 1998. Quien hubiera invertido un dólar en bonos, habría tenido una cartera valuada en quince dólares. El enigma es por qué esta discrepancia no ha provocado un alza en el valor de las accione para acercar el retorno de éstas al retorno de los bonos.
11. *Efecto de las alternativas irrelevantes*. Si cada una de dos opciones A y B es superior a la otra en una de dos dimensiones relevantes, es posible que la gente tropiece con dificultades para elegir y decida, en cambio, recoger más información sobre las opciones. Si se introduce una tercera opció C, que es 1) inferior a A en ambas dimensiones, y 2) inferior a B en una dimensión y superior en otra, hay una tendencia a elegir A sin más búsquedas.
12. *El enigma del agua fría*. En un experimento, los sujetos a quienes se había hecho creer que la cantidad de tiempo que podían mantener los brazos sumergidos en el agua extremadamente fría e

el mejor indicador de la longevidad, los mantuvieron en ella más tiempo que quienes no habían recibido esta (falsa) información.

13. *El efecto de certidumbre* (Capítulo 7). En experimentos, una mayoría prefiere ganar con certeza u viaje de una semana por Inglaterra a tener un 50% de posibilidades de ganar un viaje de tres semanas por Inglaterra, Francia e Italia, pero una mayoría también prefiere una posibilidad del 5% de la segunda opción a una posibilidad del 10% de la primera.
14. *El efecto de disyunción*. Si los sujetos de un experimento esperan ganar en una apuesta futura y se les pregunta si estarán de acuerdo en participar en una apuesta más, tienden a decir que sí. Si esperan perder, es probable que declaren la misma intención. Si no saben si van a ganar o perder, menos probable que lo hagan. El mismo efecto se observa en los dilemas del prisionero de caso único: una persona se sentirá más inclinada a cooperar si sabe que el otro ha cooperado que si sab que ha desertado y, aquí está el efecto de disyunción, *se inclinará aún más* a hacerlo si ignora la elección del otro.
15. *La maldición del ganador*. En este experimento, se pide a los sujetos que hagan una oferta por un parcela de terreno y se les dice que el vendedor conoce su valor exacto, mientras que ellos sólo saben que éste se sitúa dentro de cierta gama, en la cual todos los valores numéricos son igualmer probables. También se dice a los compradores que, de adquirirla, la parcela valdrá para ellos un 50% más que para el vendedor, porque podrán explotarla con mayor eficiencia. Si se acepta una oferta, los compradores racionales deberían poder inferir *de ese hecho* que el valor esperado de la tierra para ellos es menos de lo que han ofrecido. Si los valores oscilan entre cero y mil y se acept (digamos) una oferta de seiscientos, el comprador puede inferir que el valor real para el vendedor está entre cero y seiscientos, con un valor esperado de trescientos. Por lo tanto, su valor esperado para el comprador sería de cuatrocientos cincuenta, que es menos de lo que él ofrecía pagar. Com el mismo argumento puede plantearse para cualquier oferta que sea aceptada, los compradores racionales nunca deberían hacer una oferta. Sin embargo, en experimentos (inspirados en casos reales) nadie omite hacerla.
16. *La venganza precipitada*. Un proverbio reza que «la venganza es un plato que se sirve frío». Otro dice que «cuanto más se demora la venganza, más duro es el golpe». Es de presumir que ambos surgieron como reacción a la venganza impulsiva, y atestiguan por ello la existencia de ese fenómeno.
17. *Desdén por la eficiencia*. Montaigne escribió que «la honra de los duelos reside en el celo del val no de la ciencia; y por ello he visto a alguno de mis amigos, reputado como gran maestro en ese ejercicio, elegir en sus peleas armas que le privasen de esta ventaja y que dependieran totalmente la fortuna y de la firmeza, para que no atribuyesen su victoria a su esgrima antes que a su valor».
18. *Antes de que te cases, mira lo que haces*. Este dicho se aplica no sólo al matrimonio en sentido literal. Cuando la gente se enamora de una casa, en ocasiones muestra tanta prisa por firmar el contrato que es incapaz de descubrir defectos ocultos que se revelan más adelante.

Alternativas a la teoría de la elección racional

Para explicar estos enigmas, contamos hoy con un amplio repertorio de alternativas a la explicación basada en la elección racional. Antes de examinarlas una por una, querría enumerar los mecanismos claves en las explicaciones alternativas que consideraré (con los números de los enigmas correspondientes entre paréntesis). Algunos enigmas aparecen más de una vez, porque es posible que puedan explicarse de más de una forma.

- Aversión a la pérdida (3, 5, 7, 10).
- Ponderación no probabilística de los resultados (13).
- Descuento hiperbólico (4, 8).
- Heurística (2, 6).
- Pensamiento desiderativo (9, 12).
- Incapacidad para proyectar (15).
- Deseo de actuar por una razón (11, 14).
- Pensamiento mágico (1, 12, 14).
- Imperativo categórico (1).
- Emociones (3, 7, 14, 18).
- Normas sociales (1, 3, 16, 17).

En el pensamiento corriente, los mecanismos más preponderantes son probablemente la aversión a la pérdida y el descuento hiperbólico. En mi opinión, las emociones son una fuente aún más importante de comportamiento irracional, ya actúen en forma directa o por intermedio de las normas sociales. Aunque las emociones pueden provocar distintas perturbaciones en la racionalidad, la más significativa consiste tal vez en inducir una sensación de urgencia.

La **AVERSIÓN A LA PÉRDIDA** se define con respecto a un punto de referencia, con el supuesto de que la gente atribuye valor a los cambios desde un punto inicial determinado, y no a los estados finales existentes luego de ellos. Por lo común se considera que el punto de referencia es el statu quo, aunque es posible inducir a los sujetos a elegir otros. La aversión a la pérdida es la tendencia a atribuir mayor valor (en términos absolutos) a una pérdida desde el nivel de referencia que a una ganancia de las mismas dimensiones.[53] Desde un punto de vista empírico, se comprueba que la proporción es de alrededor de 2,5 a 1, y ésta es la que supongo en lo que sigue. Otra propiedad importante de la función de valor radica en que es cóncava para las ganancias y convexa para las pérdidas, lo cual significa que cada unidad extra de ganancia se valora menos que la anterior, y cada unidad extra de pérdida es menos dolorosa que la previa.

Dos de los enigmas pueden explicarse por el simple hecho de que las pérdidas cobran mayor importancia que las ganancias. Para resolver el enigma del corte del césped, sólo necesitamos señalar que la aversión a la pérdida predice que los costes de oportunidad y los gastos en efectivo se valoran de manera muy diferente. Como para el dueño de casa el valor de una ganancia de veinte dólares es equivalente al valor de una pérdida de ocho dólares, prefiere renunciar a la ganancia a pagar doce dólares de su bolsillo. El mismo razonamiento podría explicar el enigma de la tarjeta de crédito.

La resolución del enigma de la prima de los activos bursátiles requiere una premisa adicional, a saber, que la gente elija su combinación de bonos y acciones dentro de un horizonte temporal breve. Como el retorno de las acciones es volátil, mientras que los bonos rinden un ingreso constante año tras año, podemos ver la posesión de las primeras como la aceptación de una apuesta riesgosa. Supongamos que ofrecemos a una persona una apuesta en acciones que le da una posibilidad del 50% de ganar doscientos dólares y una posibilidad del 50% de perder cien, con el retorno fijo de los bonos como punto de referencia. Si tomamos como un hecho la aversión a la pérdida, la situación se enuncia diciendo que el valor del dinero es igual a x para $x > 0$, e igual a $2{,}5x$ para $x < 0$. Como el valor de una pérdida de cien dólares es igual (en términos absolutos) a una ganancia de doscientos cincuenta, la perspectiva de una ganancia de doscientos no puede compensar a la persona en cuestión por la perspectiva igualmente probable de una pérdida de cien. En consecuencia, la persona rechazará el ofrecimiento. Supongamos ahora que le ofrecemos un paquete de dos apuestas, a realizarse en períodos sucesivos. Esta jugada compuesta equivale a una posibilidad del 25% de ganar cuatrocientos dólares, una posibilidad del 50% de ganar cien y una posibilidad del 25% de perder doscientos. Si multiplicamos la pérdida por 2,5 para hacerla comparable a las ganancias y calculamos el valor esperado,[54] se muestra con facilidad que es 25. En consecuencia, la persona aceptará la jugada compuesta. Estudios empíricos sugieren que los inversores tienden en efecto a revalorizar sus carteras con excesiva frecuencia, una práctica miope que los induce a invertir demasiado poco en acciones y demasiado en bonos.[55]

La resolución del enigma del coste hundido sólo apela a la curvatura (convexidad o concavidad) de la función de valor. Consideremos el siguiente ejemplo. Una familia paga $\$ p$ las entradas a un partido que va a jugarse a cien kilómetros de distancia. El día del encuentro hay una tormenta de nieve. Los miembros de la familia deciden ir de todos modos, pero señalan de pasada que si les hubieran regalado las entradas, tal vez se habrían quedado en casa. Si escribimos v la función de valor para las ganancias, el valor de ir al partido es $v(g)$. Si escribimos v^* la función de valor para las pérdidas, el valor de perder $\$ p$ es el número negativo $v^*(-p)$. El coste de soportar la tormenta de nieve es c. Suponemos que $v(g) = = -v^*(-c)$, lo cual da a entender que si la familia hubiera recibido las entradas gratis le habría dado lo mismo quedarse en casa o ir al partido en medio de una tormenta de nieve. Pero como ya han pagado $\$ p$, prefieren ir. Para verlo, adviértase ante todo que debido a la convexidad de v^*, $v^*(-(c+p)) > v^*(-c) + v^*(-p)$.[56] Esto puede reescribirse como $v^*(-(c+$

$+ p)) - v^*(- c) > v^*(- p)$, que, de acuerdo con el supuesto recién enunciado, es equivalente a $v^*(- (c + p)) + v(g) > v^*(- p)$. Como el término de la izquierda en la última desigualdad es la ganancia o pérdida netas de ir al partido, y el término de la derecha es la pérdida de no ir, la familia prefiere ir.

PONDERACIÓN NO PROBABILÍSTICA DE RESULTADOS. La aversión a la pérdida se deduce de una influyente alternativa a la teoría de la elección racional llamada *teoría de la perspectiva*. Otra implicación de esa teoría es que la gente tiende a ponderar los resultados de manera diferente de lo que afirma la teoría de la utilidad esperada. De acuerdo con esta última teoría, la utilidad es de probabilidades lineales (Capítulo 11). La teoría de la perspectiva, en contraste, sostiene que la gente es más sensible a los cambios de probabilidad cerca de los límites naturales de 0 (imposible) y 1 (seguro). El efecto de certidumbre ilustra la no linealidad en torno de 1. Los creadores de la teoría de la perspectiva, Daniel Kahneman y Amos Tversky, citan el siguiente ejemplo (que atribuyen a Richard Zeckhauser) de la no linealidad en torno de 0:

> Supongamos que usted se ve obligado a jugar a la ruleta rusa, pero le dan la oportunidad de comprar la eliminación de una bala del revólver cargado. ¿Pagaría lo mismo para reducir la cantidad de balas de cuatro a tres que para reducirla de uno a cero? La mayoría de la gente siente que estaría dispuesta a pagar mucho más por la reducción de la muerte de un sexto a cero que por una reducción de cuatro sextos a tres sextos. Las consideraciones económicas [es decir la teoría de la utilidad esperada] nos llevarían a pagar más en el último caso, donde el valor del dinero queda presuntamente reducido a causa de la considerable probabilidad de que no vivamos para disfrutarlo.

El **DESCUENTO HIPERBÓLICO** se examinó en el Capítulo 6. Aquí me gustaría limitarme a señalar el estrecho vínculo entre los enigmas 8 y 4. Si la gente participa de los clubes navideños, lo hace presumiblemente porque sabe que si deposita sus ahorros en una cuenta normal con la intención de guardarlos allí hasta la Navidad, cambiará de opinión y volverá a retirarlos.

HEURÍSTICA. La heurística (reglas prácticas) puede despistarnos. La creencia del jugador de que la ruleta tiene una memoria puede surgir o bien de la heurística de la representatividad («es hora de que salga el rojo») o bien de la heurística de la disponibilidad («el rojo está en racha»). La preferencia por el ruido más desagradable deriva del uso de una heurística «de máximo y final» [*«peak-end»*], de conformidad con la cual las experiencias pasadas se valoran según se vivieron en su mejor (o peor) nivel y en el final, no en función de su calidad general de agradables o desagradables. Esta heurística llevaría a las personas a preferir la peor experiencia en términos objetivos, dado que tiene un mejor final que la otra y el mismo máximo.

PENSAMIENTO DESIDERATIVO. El fenómeno del pensamiento desiderativo se analizó en el Capítulo 7. Este fenómeno puede ser provocado por un simple deseo, como cuando los individuos que se desempeñan en ocupaciones riesgosas bien pagas subestiman los riesgos que corren. Es aún más probable que ocurra cuando el deseo surge de una emoción fuerte, como es el caso de los pacientes cancerosos terminales que eligen un tratamiento cuyo único efecto es hacerlos sufrir más. Es posible que el pensamiento desiderativo inducido por el deseo de tener una larga vida también ejerza su influencia en el enigma del agua fría.

INCAPACIDAD PARA PROYECTAR. En numerosas situaciones, la gente toma malas decisiones debido a su incapacidad para proyectarse en el futuro. Con esto me refiero a la falta de capacidad para imaginar lo que uno mismo u otros tendrían razones para creer, o incentivos para hacer, en situaciones futuras que dependen de la elección presente. La maldición del ganador puede explicarse en función de esa incapacidad. Otro ejemplo es el del presidente Chirac y su desastrosa convocatoria a elecciones anticipadas en junio de 1997. La derrota de su coalición tal vez se haya debido a que los votantes interpretaron que, si Chirac quería elecciones anticipadas, era porque sabía algo que ellos desconocían y que lo llevaba a creer que, de esperar, perdería. Al convocar esas elecciones, el presidente reveló lo que sabía o, al menos, reveló *que* sabía algo desfavorable y, por consiguiente, dio al electorado una razón para votar contra él. Las encuestas le decían que ganaría, pero *las encuestas difieren de las elecciones*, dado que el hecho de realizarlas no revela a los participantes nada sobre las creencias de la persona que las ha encargado.[57]

DESEO DE ACTUAR POR UNA RAZÓN. Mencioné varios ejemplos de este mecanismo al comienzo del presente capítulo. En el enigma 11, el deseo lleva al agente a modificar su comportamiento cuando se le presenta una opción que es indudablemente inferior a una de las opciones con que ya cuenta.[58] En otros casos, el agregado de opciones puede impedir al agente tomar cualquier tipo de decisión. Comportamientos similares al del asno de Buridan, que murió de hambre porque no podía decidir cuál de dos montones idénticos de heno comer, han sido confirmados en ámbitos de la vida real. Un psicólogo que instaló puestos de venta de mermelada en Broadway comprobó que, cuando tenían una amplia variedad de marcas, los transeúntes los miraban más pero compraban menos, en comparación con los puestos provistos de menos variedades. Cuando hay más opciones, es más difícil decirse sin vacilar: «ésta

es la mejor». Quienes necesiten basar su elección en razones suficientes se abstendrán de elegir.

Tal como lo sugiere el enigma 14, para actuar por una razón es necesario *tenerla*, no meramente *saber* que la tenemos. Así, supongamos que sé exactamente que suceden *p* o *q*, pero no sé cuál. Si sucede *p*, tengo una razón para hacer X. Si sucede *q*, también tengo una razón para hacer X. Por lo tanto, sé que cualquiera sea el caso, tengo una razón para hacer X, pero como no sé *qué* razón, me abstengo de hacerlo. El enigma 14 propone un ejemplo de esta anomalía.[59] Para considerar otro ejemplo, tomemos el hecho de que en el antiguo derecho inglés un acusado debía ser absuelto si las pruebas dejaban dudas acerca de si había cometido un robo o un desfalco, a pesar de que, de probarse uno de los dos cargos, habría sido condenado. Para serlo, tendría que habérselo declarado culpable o bien de *p* o bien de *q*. El hecho de que se lo declarara culpable de (*p* o *q*) no era suficiente. Otra respuesta al aprieto podría consistir en postergar la decisión. Supongamos que el Gobierno de los Estados Unidos debate si establecer o no una relación comercial con otro país en el cual habrá próximamente elecciones presidenciales. Si gana el candidato proestadounidense, los Estados Unidos tendrán razones suficientes para establecer esas relaciones. Si triunfa el candidato antiestadounidense, los Estados Unidos tendrán (otro conjunto de) razones suficientes para implementar la misma política. No obstante, podrían esperar hasta después de los comicios para anunciar su decisión. En ciertas circunstancias, la demora podría ser no sólo costosa sino innecesaria.

PENSAMIENTO MÁGICO. El mecanismo del pensamiento mágico (Capítulo 7) podría explicar el comportamiento en el enigma del agua fría. Quizá pueda explicar también algunos casos del efecto de disyunción. Si en el dilema del prisionero la gente se inclina más a cooperar cuando no sabe si el otro participante ha cooperado o desertado, esa actitud tal vez se deba a que cree, mágicamente, que al cooperar puede provocar la cooperación del otro. «Al ser como yo, actuará como yo.» Las intenciones de voto también pueden formarse de este modo. Si creo, irracionalmente, que mi voto no se limita a ser un predictor del voto de otros, sino que de alguna manera hace más probable que ellos voten, la mayor eficacia de mi acción da una apariencia racional al hecho de votar.

IMPERATIVO CATEGÓRICO. Hay una estrecha relación entre esta última instancia de pensamiento mágico, el imperativo categórico y (una versión

cotidiana de) el imperativo categórico de Kant, según el cual uno debe hacer A y no B si se da el caso de que, al hacer todos A y no B, todos estaremos mejor. Sin embargo, la actuación sobre la base del imperativo categórico es irracional. La racionalidad me indica elegir en función de lo que sucederá si *yo* hago A y no B.[60] El imperativo categórico me indica elegir en función de lo que sucederá si *todo el mundo* hace A y no B. En unas elecciones nacionales, aun quienes no están sujetos al pensamiento mágico podrían «abstenerse de abstenerse» con la idea de que «¿qué pasaría si todo el mundo hiciera eso?»

EMOCIONES. Para comparar el comportamiento basado en las emociones con el comportamiento racional, podemos modificar la Figura 11.1 para incluir (en las líneas más marcadas) el impacto de la emoción en cada uno de los elementos del esquema (véase la Figura 12.1)

Se ha sostenido que las emociones pueden afectar directamente la *acción*, en los casos de debilidad de la voluntad (Capítulo 6). Medea, al matar a sus hijos para vengarse de Jasón, sabe, *mientras los mata*, que actúa contra su parecer. Dejé asentadas mis reservas con respecto a la idea, pero no puede excluírsela. Las emociones afectan de dos maneras los *deseos*. Primero, en virtud de la tendencia a la acción asociada, tal vez generen un cambio temporario de preferencias. Si la situación impone una demora entre el momento en que se toma la decisión de actuar y el momento de actuar, es posible que la acción nunca se lleve a cabo. Un ejemplo mencionado en el Capítulo 8 era el hecho de que las mayores expresiones de interés en servir en el ejército, luego del 11 de septiembre de 2001, no condujeron al aumento de la cantidad de personas reclutadas. Si la acción puede emprenderse de inmediato, a veces quizá se revierta más adelante, al decrecer la emoción. Así, entre los doscientos mil hombres que desertaron del ejército de la Unión durante la Guerra de Secesión de los Estados Unidos de América, presuntamente por miedo, el 10% regresó de manera voluntaria. En algunos casos, sin embargo, la acción llevada a cabo sobre la base de preferencias temporarias inducidas por la emoción es irreversible (una situación de trampa de langostas). Tras incorporarse a un movimiento guerrillero en un momento de entusiasmo, los hombres y mujeres jóvenes que luego quieren abandonarlo pueden llegar a comprobar que esta opción no está a su alcance. Segundo, las emociones pueden provocar un aumento temporario del índice de descuento temporal, y hacer de ese modo que opciones antes no preferidas, con malas consecuencias de largo plazo, parezcan preferibles.

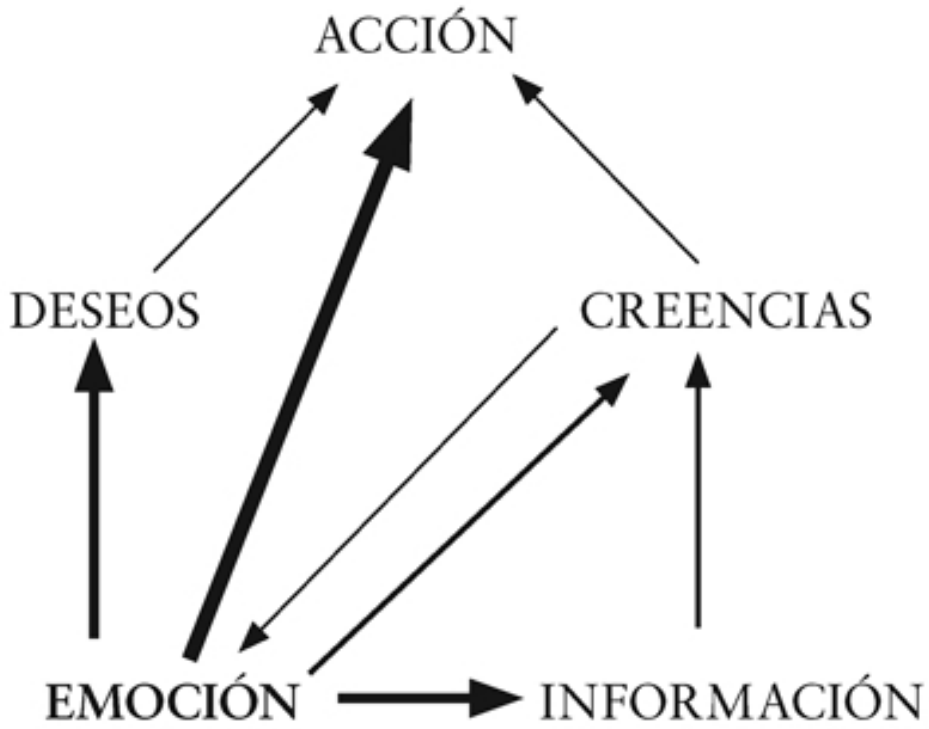


FIGURA 12.1

Las emociones pueden afectar directamente las *creencias* a través de los mecanismos del pensamiento desiderativo y el pensamiento contradesiderativo. La emoción de la altivez, basada en la creencia de que uno es un tipo superior de persona, resistirá la admisión de que uno ha cometido un error. Esto puede explicar, al menos en algunos casos, la vulnerabilidad de la falacia del coste hundido.[61] La altivez quizá también explique por qué las personas tienden a culpar a sus víctimas:[62] como la autoestima no les permite admitir la culpa, encuentran en el otro una falla que no sólo justifica lo que le han hecho, sino que motiva nuevos comportamientos agraviantes. Al igual que en la falacia del coste hundido, la falta de voluntad para reconocer un error pasado puede generar peores consecuencias que el error mismo. En cuanto a las emociones y el pensamiento contradesiderativo, veremos en el Capítulo 23 varios ejemplos de pánico inducido por el miedo.

Por último, en virtud de su carácter de urgencia (Capítulo 8), las emociones pueden interferir en la adquisición óptima de *información* y, en consecuencia, también afectar indirectamente las creencias. La ira (enigma 16) y el amor (18) nos llevan a hacer cosas que no habríamos hecho de haber

tenido una política más racional de recolección de información. Una observación de Séneca quizá nos ayude a reunir algunos de estos hilos: «La ira carece por completo de equilibrio; ora va más lejos de lo que debe, ora se detiene antes de lo que debería». A menudo, la urgencia induce un descuido de los efectos temporalmente remotos de las opciones, por obra del hecho de que *la determinación de las consecuencias de largo plazo consume tiempo de por sí*. Por ende, no es imperioso que el truncamiento del horizonte temporal se deba a un descuento más alto de las consecuencias conocidas; puede deberse al hecho de que algunas consecuencias ni siquiera aparecen en la pantalla mental del agente.

NORMAS SOCIALES. Las emociones del desprecio y la vergüenza tienen un importante papel en el mantenimiento de las normas sociales (Capítulo 21). La precipitación vengativa puede deberse a la urgencia de la ira, pero también podría tener su origen en una norma social que estigmatiza como cobarde a quienquiera que se demore en vengarse. La negativa de los duelistas a escoger las armas para las que tienen mayor aptitud también es sostenida por el temor a ser vistos como excesivamente preocupados por la mera supervivencia y no por el honor. Para terminar, el enigma del corte del césped podría encontrar su explicación en el funcionamiento de normas sociales y no en la aversión a la pérdida. A un residente no se le ocurriría cortar el césped de su vecino, porque en las comunidades suburbanas hay una norma social que impide a un adulto hacer cosas semejantes por dinero. Sencillamente, *eso no se hace*. El voto puede reflejar asimismo la intervención de normas sociales si el acto de votar es visible para otros y éstos desaprueban a quienes no votan. Es más plausible ver en el hecho de concurrir a votar en grandes elecciones anónimas el resultado de normas *morales*, que pueden en sí mismas contener aspectos irracionales (véase el examen anterior).

* * *

NOTA BIBLIOGRÁFICA

Se encontrará una evaluación crítica de la teoría de la elección racional desde una perspectiva un tanto diferente de la presente en Donald P. Green e Ian Shapiro, *Pathologies of Rational Choice: A Critique of Applications in Political Science*, New Haven, Yale University Press, 1994, que puede complementarse provechosamente con Jeffrey Friedman (comp.), *The Rational Choice Controversy: Economic Models of Politics Reconsidered*,

New Haven, Yale University Press, 1996. Sobre la idea de la racionalidad como norma, véase Dagfinn Føllesdal, «The status of rationaliy assumptions in interpretation and in the explanation of action», en Michael Martin y Lee C. McIntyre (comps.), *Readings in the Philosophy of Social Science*, Cambridge (Massachusetts), The MIT Press, 1994, págs. 299-310. Sostengo el carácter hiperracional del principio del «interés superior del niño» en el Capítulo 3 de *Solomonic Judgments: Studies in the Limitations of Rationality*, Cambridge y Nueva York, Cambridge University Press, 1989 [*Juicios salomónicos: las limitaciones de la racionalidad como principio de decisión*, Barcelona, Gedisa, 1991]. En Jonathan Wiener, «Managing the iatrogenic risks of risk management», *Risk: Health, Safety and Environment*, 9, 1998, págs. 39-82, se hallará un fructífero examen de algunos problemas conexos. La mayor parte de los enigmas subsiguientes se analizan en artículos reeditados en los libros de referencia enumerados en la nota bibliográfica al Capítulo 7. Las excepciones son el enigma 1, para el cual véase André Blais, *To Vote or Not to Vote?: The Merits and Limits of Rational Choice Theory*, Pitsburgo, University of Pittsburgh Press, 2000; el enigma 2, para el cual véase Daniel Kahneman, «Objective happiness», en Daniel Kahneman, Ed Diener y Norbert Schwartz (comps.), *Well-Being: The Foundations of Hedonic Psychology*, Nueva York, Russell Sage, 1999, págs. 3-25; los enigmas 16 y 17, para los cuales véase el Capítulo 3 de mi *Alchemies of the Mind: Rationality and the Emotions*, Cambridge y Nueva York, Cambridge University Press, 1999 [*Alquimias de la mente: la racionalidad y las emociones*, Barcelona, Paidós, 2002], y el enigma 18, para el cual véase el análisis del Capítulo 8 de este libro. La mayoría de las alternativas se examinan en los mismos libros de referencia, excepto el pensamiento desiderativo (Capítulo 7), las emociones (Capítulo 8) y las normas sociales (Capítulo 21). Debería señalarse que el efecto de certidumbre tiene estrecha relación con el mismísimo primer enigma que se presentó explícitamente (en 1953) como un reto a la teoría de la elección racional, la «paradoja de Allais». El «enigma Chirac» se discute, junto con muchos ejemplos similares, en Alastair Smith, *Election Timing*, Cambridge y Nueva York, Cambridge University Press, 2004. La idea de que los animales no incurren en la falacia del coste hundido es sostenida en Hal Arkes y Peter Ayton, «The sunk cost and Concorde effects: are humans less rational than lower animals?», *Psychological Bulletin*, 125, 1999, págs. 591-600. El ejemplo de las ventas de mermeladas en Broadway se extrae de Sheena Iyengar y Mark Lepper, «When choice is demotivating: can one desire too much of a good thing?», *Journal of*

Personality and Social Psychology, 79(6), 2000, págs. 995-1006. La re referencia al desfalco y el robo en el viejo derecho inglés pertenece a James Fitzjames Stephen, *History of the Criminal Law of England*, Londres, Macmillan, 1883, y Búfalo (Nueva York), Hein, 1964, vol. 3, pág. 153. Los descubrimientos acerca de la reincorporación al ejército de la Unión provienen de Dora Costa y Matthew Kahn, «Deserters, social norms and migration», *The Journal of Law and Economics*, 50, 2007, págs. 323-353. En Dianne Tice, Ellen Bratslavsky y Roy Baumeister, «Emotional distress regulation takes precedence over impulse control: if you feel bad, do it!», *Journal of Personality and Social Psychology*, 80(1), 2001, págs. 53-67, se presentan pruebas sobre un aumento del descuento temporal inducido por la emoción.

Notas al pie

[46] Actuar por *una* razón no debe confundirse con actuar de acuerdo con la razón, tal como la idea se definió en el Capítulo 4. La persona que elige en función de su egoísmo actúa por una razón, pero no de acuerdo con la razón. A la inversa, la persona que actúa sobre la base de un imperativo categórico actúa de acuerdo con la razón, pero no por una razón (véase la nota 19 de ese capítulo). Los lectores que consideren contraintuitivo el último enunciado probablemente tengan una idea diferente de la mía de lo que significa «actuar por una razón».

[47] De hecho, el «interés superior del niño» es el criterio utilizado en la mayoría de las leyes de custodia.

[48] Como era un erudito shakesperiano, es posible que Johnson tuviera en mente la observación del rey en *Hamlet* (tercer acto, escena 3): «Como el hombre a quien dos obligaciones llaman, me detengo a considerar por cuál empezaré primero, y no cumplo ninguna».

[49] Hay dos razones por las cuales sólo puede esperarse que la igualdad sea aproximada. Primero, las conmociones externas siempre provocarán desviaciones de la igualdad. Segundo, la aversión al riesgo puede inducir valores inferiores (y por lo tanto tasas de retorno más altas) en activos sumamente volátiles.

[50] Algunos podrían estar tentados de reemplazar «o» por «y» en esta frase. Sin embargo, excepto por suerte, uno no puede maximizar dos objetivos al mismo tiempo. Para ser más preciso, el agente procuraría la mezcla factible óptima de las dos metas, según la representan un conjunto de oportunidades y una familia de curvas de indiferencia (Capítulo 9).

[51] Muchos de los ejemplos se han mencionado en capítulos anteriores y se resumen aquí por razones de conveniencia. Sus fuentes de procedencia son diversas: proverbios, autores clásicos, experimentos mentales, experimentos de laboratorio y observaciones de la vida real. Sin embargo, los ejemplos de las primeras tres categorías se basan en teorías bien establecidas que se examinan a continuación en este mismo capítulo.

[52] La paradoja se plantea cuando el único objetivo de los votantes es poner un candidato en el cargo o una propuesta en vigencia. No se planteará inevitablemente cuando la meta sea contribuir a la vitalidad del sistema democrático o dar un «mandato» a un candidato, dado que en estos casos los votantes son importantes, aunque no sean el eje de la cuestión. Sin embargo, aun cuando la motivación sea respaldar la democracia, la explicación quizá se encuentre en uno de los mecanismos no racionales que se analizan más adelante.

[53] Suponiendo que los bienes (incluido el dinero) tienen una utilidad marginal decreciente, la teoría clásica de la elección racional también predice que las pérdidas tendrán mayor peso que las ganancias de iguales dimensiones a partir del mismo punto inicial. La magnitud del efecto, sin embargo, suele ser mucho menor. Además, la teoría clásica de la utilidad da a entender que la ganancia de utilidad de moverse de A a B es igual a la pérdida de utilidad de moverse de B a A, ya que esas diferencias se deducen simplemente de la comparación de los *niveles* de utilidad de los dos estados.

[54] Como la teoría de la perspectiva supone que los pesos de las decisiones difieren de las probabilidades, este cálculo sólo es aproximadamente correcto. Ni esta simplificación ni el supuesto de una función de valor lineal importan para la conclusión del análisis.

[55] El término «miope» no tiene aquí el mismo significado que en los análisis del descuento temporal (Capítulo 6). No se refiere al modo como el agente calcula el valor presente de flujos futuros de ingresos, sino a la tendencia a tomar decisiones sucesivas por separado en vez de «agruparlas» en una elección general. Esa «miopía decisional», como podríamos llamarla, también podría actuar en otros contextos. Así, cuando la gente trata de controlar el descuento hiperbólico

«amontonando» elecciones sucesivas (Capítulo 13), su éxito tal vez dependa de la cantidad de elecciones que incluyen.

[56] Como ésta es una comparación de dos números negativos, la desigualdad enuncia que el primero está más cerca de cero, es decir que es más pequeño en términos absolutos.

[57] En la quinta parte examino el comportamiento de Chirac como un ejemplo de lo que llamo «síndrome del hermano menor».

[58] Esta formulación es un tanto engañosa, puesto que en experimentos no son los *mismos* sujetos quienes quedan expuestos en primer lugar al conjunto de elección (A, B) y luego al conjunto (A, B, C), sino dos grupos diferentes de sujetos repartidos al azar en cada uno de ellos. La interpretación natural del descubrimiento, sin embargo, es que los sujetos del grupo (A, B, C) *se habrían comportado* como los sujetos del grupo (A, B) si sólo se les hubiesen presentado las dos opciones. La «modificación», en consecuencia, se refiere a un punto inicial contrafáctico, no a uno real. Esta observación es válida para muchos de los experimentos mencionados en este libro.

[59] En términos similares, se ha demostrado que la cantidad máxima que la gente está dispuesta a pagar por una lotería entre dos opciones es menor que el máximo que está dispuesta a pagar por la menos atractiva de ellas.

[60] La racionalidad no me indica elegir en función de lo que *me* sucederá si hago A y no B (véase el Capítulo 11). Es compatible con algunas formas de moral referida a los otros, pero no con la representada por el imperativo categórico.

[61] En este contexto, es interesante que los animales no parezcan incurrir en esa falacia: tal vez es porque no tienen ninguna autoestima por la cual preocuparse.

[62] Como señalo hacia el final del Capítulo 15, hay asimismo una tendencia a culpar a las víctimas de una acción perpetrada por un tercero, como ocurre cuando la gente, en vez de culpar al violador, culpa a su víctima. En este caso, el mecanismo no es la altivez sino una creencia o un supuesto tácito en el sentido de que, puesto que el mundo es fundamentalmente justo, la víctima debe haber merecido de algún modo su destino.

Capítulo 13

Respuestas a la irracionalidad

Racionalidad secundaria

En los dos últimos capítulos hemos considerado el ideal del comportamiento racional y los frecuentes traspiés de la racionalidad. Por difundidos y frecuentes que sean, esos traspiés no son inevitables. Si entendemos nuestra propensión a cometer errores, podemos tomar y tomamos efectivamente precauciones para disminuir la probabilidad de volver a cometerlos, o al menos para limitar el daño en caso de hacerlo. Como he dicho reiteradas veces, *queremos* ser racionales. Quizá concibamos esas estrategias precautorias como una forma de *racionalidad* imperfecta o *secundaria*. Es preciso distinguirlas del mero *aprendizaje*, que ocurre cuando la propensión sencillamente se borra como resultado de una perspicacia mejorada. Se ha informado, por ejemplo, que cuando la gente se da cuenta de que votar es en cierto sentido inútil, se inclina menos a hacerlo.[63] Las falacias cognitivas que son afines a ilusiones ópticas también pueden superarse por medio del aprendizaje. Así como todos aprendemos a ignorar la apariencia de un palo que parece quebrado en el agua, algunos jugadores aprenden, presumiblemente de la peor manera, que los dados no tienen memoria. Sin embargo, lo que me interesa aquí son las propensiones que persisten a lo largo del tiempo.

Para hacer frente a nuestras tendencias a comportarnos irracionalmente, es posible que apelemos o bien a estrategias intrapsíquicas o bien a dispositivos extrapsíquicos (*precompromiso*). Presentaré ante todo una

ilustración del uso de estas técnicas para contrarrestar el descuento hiperbólico y el comportamiento inconsistente generado por éste, y luego examinaré su utilización con fines de control del comportamiento emocional y adictivo. Si bien diversas estrategias no son necesariamente racionales, muchas lo son.

Los selves futuros como aliados

Un agente que está sujeto al descuento hiperbólico y lo sabe, es *sofisticado*. A diferencia del agente ingenuo que cambia una y otra vez de opinión sin entender el mecanismo subyacente, el agente sofisticado es consciente de su propensión y la deplora. En previsión de situaciones futuras en las cuales se verá frente a la elección entre una pequeña recompensa anticipada y una recompensa demorada más grande, le gustaría obligarse a elegir la última a pesar de su inclinación a escoger la primera. En algunos casos, puede tratar sus «selves futuros» como aliados en un esfuerzo común por superar las tentaciones. En otros, quizá los trate como adversarios y procure limitar el daño que pueden hacer a su «yo presente». Este vocabulario es metafórico, por supuesto, pero lo desmetaforizaremos.

Consideremos en primer lugar el caso en que cabe esperar que la elección entre una pequeña recompensa anticipada y una gran recompensa demorada se plantee una y otra vez. El agente puede obligarse entonces a buscar la recompensa demorada si *agrupa* (o *lía*) las elecciones.

Querría utilizar como ilustración un ejemplo de la época en que yo vivía en las colinas cercanas a la universidad donde dictaba clases. Todos los días usaba la bicicleta para llegar al campus y volver a casa. El viaje de regreso implicaba algunas subidas bastante empinadas, de modo que yo enfrentaba a diario la tentación de bajarme de la bicicleta y caminar, en vez de obligarme a pedalear. Cuando salía del campus me hacía el firme compromiso de permanecer en la bicicleta todo el camino, pero en medio de la trepada solía ocurrírseme una idea seductora: «¿Por qué no caminar hoy, y volver a pedalear mañana?» Luego, por fortuna, aparecía otra idea: «¿Qué hay de especial en mañana? Si hoy cedo a la tentación, ¿no predice esa actitud que también lo haré mañana, pasado mañana, etc.?» Este último pensamiento me permitía mantenerme montado en la bicicleta.

Este dispositivo intrapsíquico entraña una *reformulación* de la situación. En vez de concebir que los futuros viajes de vuelta a casa implicaban una

serie de elecciones, comencé a verlos como *una elección entre dos series*: subir siempre la colina montado en la bicicleta y caminar siempre con ésta al lado. Al decirme que mi comportamiento en una ocasión era lo que mejor predecía mi comportamiento en la ocasión siguiente, yo establecía un efecto interno de dominó que elevaba las apuestas y me hacía inclinarme por la recompensa demorada de una mejor salud, y no por la recompensa anticipada del alivio de la incomodidad. Con referencia a la Figura 6.1, puede mostrarse, de hecho, que si ponemos muchos pares de recompensas idénticas a A y B en la línea horizontal y luego creamos dos curvas, una para la suma de las curvas de valores presentes de todas las pequeñas recompensas y otra para la suma de las curvas de valores presentes de todas las grandes recompensas, la última curva se situará encima de la primera en el momento en que tiene que hacerse la primera elección, siempre que la cantidad de elecciones sucesivas que deben hacerse sea lo bastante grande.[64] En otras palabras, el agrupamiento de las elecciones puede hacer que la opción de inclinarse siempre por la recompensa más grande sea preferible a la de decidirse siempre por la recompensa más pequeña. Está claro que la elección de la recompensa más pequeña hoy y la recompensa más grande en todas las futuras ocasiones es aun mejor, pero según el supuesto planteado esta opción no figura en el conjunto de oportunidades del agente.[65]

¿Puede justificarse ese supuesto? ¿*Es* mi comportamiento de hoy un buen predictor de mi comportamiento de mañana? En los casos en que hay un auténtico efecto causal, tal vez sea así. Mi viaje en bicicleta de hoy contribuirá a la fortaleza de mis músculos, de modo tal que mañana también pueda andar en ella.[66] En mi caso, sin embargo, me he apoyado en el pensamiento mágico y no en la eficacia causal. Así como mucha gente vota o hace donaciones a entidades benéficas bajo la influencia de la idea «si yo no, ¿quién?», lo que me mantenía en la bicicleta era la idea «si no ahora, ¿cuándo?» O, para decirlo de manera más elaborada: «El día de hoy no tiene nada de especial. Si bajo de la bicicleta, las causas que me han llevado a hacerlo también actuarán mañana e inducirán el mismo comportamiento. Si no hago un esfuerzo ahora, no lo haré nunca». Con todo, en ausencia de un auténtico efecto causal, no se deduce esa conclusión. Si hoy *puedo* seguir andando en bicicleta pero decido bajarme, también puedo seguir andando mañana. Aunque falso, el razonamiento tiene peso y, creo, suma difusión. Y muestra que podemos adoptar una forma de irracionalidad (el pensamiento mágico) para combatir otra (el descuento hiperbólico).[67]

Para funcionar bien, esas estrategias quizá deban formularse como elecciones binarias: hacerlo siempre o nunca. Para mucha gente, la abstención es más fácil que la moderación. Boswell señaló que «Johnson, aunque podía ser estrictamente *abstemio*, no era un hombre *contenido* ni en el comer ni en el beber. Podía refrenarse, pero no consumir moderadamente». El mismo problema se plantea si, en vez de limitar el consumo en cada ocasión, uno trata de limitar la cantidad de ocasiones en las que puede abandonarse. La estratagema de establecer de antemano qué se entenderá por ocasión legítima es fácil de invalidar. Tal vez la gente resuelva no beber alcohol antes de cenar, sólo para encontrarse con que programa las cenas cada vez más temprano. La regla de beber vino únicamente en los restaurantes, y nunca en casa, puede impulsarnos a salir a cenar con más frecuencia. La regla kantiana de fumar sólo una pipa después del desayuno (Capítulo 4) no era lo bastante inequívoca para brindar protección a Kant, dado que con el paso del tiempo éste tomó el hábito de comprar pipas cada vez más grandes. Cuando es factible, la regla «no hacerlo nunca» es acaso la única capaz de sostenerse de manera estable. Como este curso de acción no es factible con respecto a la comida, la obesidad puede ser más recalcitrante a las reglas privadas que las adicciones.

La formulación de elecciones binarias puede, sin embargo, inducir un comportamiento absurdamente rígido. Supongamos que me he dicho que no he de tolerar una sola excepción a la regla de cepillarme los dientes todas las noches. En una ocasión determinada, me encuentro sin un cepillo de dientes y decido caminar ocho kilómetros en medio de una ventisca para comprar uno. Para sostener la decisión, me digo que si rompo la regla en esa oportunidad, sentaré un precedente peligroso que me llevará a violarla por razones cada vez más triviales, de tal manera que pronto no habrá regla en absoluto y se me caerán todos los dientes. Algunas personas construyen sistemas muy elaborados de este tipo, en los que la omisión de la observancia de una regla predice la inobservancia de otras, lo cual eleva aún más las apuestas.[68] Como las reglas privadas pueden tener estos efectos frustrantes, el remedio que proporcionan es a veces peor que la enfermedad. En lenguaje freudiano (Capítulo 4), el estricto control de los impulsos ejercido por el superyó podría ser más perjudicial que los impulsos del ello.

Los futuros *selves* como adversarios

Consideremos ahora el caso en que el agente enfrenta una elección entre recompensas (o castigos) en una de varias fechas futuras. (A diferencia del caso anterior, se supone que la elección se plantea una sola vez.) El agente puede valerse entonces del dispositivo intrapsíquico consistente en responder estratégicamente a la propensión conocida de los «yos futuros» a descontar el futuro de manera hiperbólica. Imaginemos que soy un «postergador hiperbólico» que siempre deja las cosas para mañana y luego, cuando llega ese día, las pospone otra vez hasta el día siguiente. Una vez que comprendo que estoy sujeto a esa propensión, mi comportamiento óptimo cambia. Supongamos que puedo llevar a cabo determinada tarea desagradable en cualquiera de tres plazos, y que el coste de hacerlo sube con el tiempo. Si soy ingenuo, quizá me diga que realizaré la tarea mañana. Si soy sofisticado, sé que mañana la pospondré hasta el último plazo. La comprensión de que el coste puede, en realidad, ser muy alto a menos que cumpla la tarea ya mismo, tal vez me induzca a hacer exactamente eso.[69]

En este caso, es una ayuda ser sofisticado. En otros, tal vez sea mejor ser ingenuo. Supongamos que podemos obtener una recompensa en cualquiera de tres plazos sucesivos y que dicha recompensa aumenta con el tiempo. Un ejemplo podría ser el de una persona a quien le han ofrecido una botella de vino que mejora con el tiempo hasta el tercer año, y luego se deteriora. Una persona ingenua quizá se forme la intención de esperar hasta el tercer plazo, y luego cambie de opinión y beba el vino en el segundo. Un individuo sofisticado sabrá que nunca va a esperar hasta el tercer plazo, de modo que, en concreto, sólo enfrenta la elección entre la recompensa del primer plazo y la del segundo. En esa elección, es posible que se imponga la recompensa anticipada.[70] Algunos alcohólicos dicen estar sujetos a un tipo similar de razonamiento: «Sé que voy a ceder a la tentación, así que bien podría hacerlo ya mismo». Del mismo modo, los fumadores ingenuos que abandonan el hábito por primera vez quizás aguanten más tiempo que fumadores sofisticados que lo han intentado varias veces y conocen las escasas probabilidades de éxito. Aunque la recaída en la adicción no tiene por qué deberse a un descuento hiperbólico, el sentido general es el mismo: si podemos predecir que nos desviaremos de nuestro mejor plan, acaso terminemos por desviarnos de él aún más o antes que si ignoráramos que vamos a fracasar.

Dispositivos extrapsíquicos

En la práctica, la planificación sofisticada contra nuestros *selves* futuros es probablemente menos importante que las estrategias de agrupamiento y los dispositivos de precompromiso de los que voy a ocuparme ahora. Éstos implican afectar el mundo externo, de una manera que no puede deshacerse al instante y sin costes, con el propósito de reducir la probabilidad de que elijamos en el futuro la recompensa anterior y más pequeña. Destacan cinco estrategias: *eliminar* del conjunto factible la elección de la recompensa anticipada; *imponer una sanción* a la elección de la recompensa anticipada; *agregar un premio* a la elección de la recompensa demorada; *imponer una demora* entre la elección y la entrega concreta de la recompensa, y *evitar señales* que puedan provocar una inversión de preferencias. El comportamiento ahorrativo puede ilustrar las primeras cuatro opciones. Si comienzo a ahorrar para Navidad pero compruebo que saco dinero de mis ahorros en vez de guardarlo, tal vez me incorpore a un club de ahorros navideño que no permita retiros anticipados (Capítulo 12). O bien puedo depositar mis ahorros en una cuenta que dé mayor interés pero imponga una penalidad a los retiros anticipados, y combine con ello premios y castigos. Si quiero ahorrar para la vejez, puedo establecer un plazo entre una decisión de desahorrar y el momento de tener los fondos a mi disposición, invirtiendo en activos ilíquidos y no en acciones o bonos. La quinta opción tiene su ilustración en la persona cuyo antojo de postre es provocado por señales visuales. El truco consiste en no ir a un restaurante donde los postres circulen en una mesa rodante, de modo que haya que elegirlos en el menú. Podemos contrastar esta situación con la del individuo que tiene un problema con los postres debido al descuento hiperbólico. En su caso, la mejor opción es ir a un restaurante donde tenga que pedir el postre al comienzo de la comida.

Las personas que se inscriben en un gimnasio para hacer ejercicios semanales a menudo abandonan al cabo de una o dos semanas. Para impedirlo, pueden (al menos en teoría) firmar un contrato con el gimnasio por el cual se comprometen a pagar por anticipado el doble del arancel normal y recibir un reembolso parcial cada vez que se presentan. Las personas que se inscriben en programas de pérdida de peso tal vez tengan que pagar un depósito que sólo recuperarán si rebajan una cantidad determinada de kilos, a veces con la cláusula de que, si no lo logran, el

depósito será donado a la causa política menos apreciada por ellas. Cuando me propuse dictar las conferencias que resultaron en el presente libro, me comprometí previamente al decir a mis estudiantes que les daría el borrador de un capítulo al final de cada semana. Si no hubiera conseguido estar a la altura de esa promesa, habría sufrido el coste de un moderado ridículo. Si me temo que voy a cancelar mi cita con el dentista cuando la fecha se acerque, puedo autorizarlo a facturarme el doble de los honorarios regulares en caso de no aparecer por su consultorio. En lo concerniente al vino que mejorará con el tiempo, podemos pedirle al vendedor que nos lo guarde para protegernos de una gratificación prematura. Si nos inquieta la posibilidad de leer demasiado rápidamente la última novela de nuestro escritor favorito de policiales, salteándonos párrafos para llegar al desenlace, podríamos comprarnos una versión grabada del libro (y un grabador que no tenga la función de adelantar la cinta) que no nos deje otra opción que escuchar todas las palabras.

Los ejemplos de los dos últimos párrafos encierran un precompromiso contra dos tipos de tentación. Por un lado está la *dilación*, incluida la incapacidad de ahorrar, buscar tratamientos para el dolor, hacer ejercicios físicos o redactar un manuscrito. Por otro, está la *satisfacción prematura*, demostrada, por ejemplo, cuando se bebe un vino demasiado joven o se saltean páginas de un libro. Estas tentaciones se originan directamente en el descuento hiperbólico. No interviene otra cosa que el mero transcurso del tiempo. En otra categoría de casos, el *comportamiento excesivo*, el descuento hiperbólico puede interactuar con otras motivaciones, de carácter visceral. Entre ellas se incluyen el exceso de comida, el juego compulsivo y el comportamiento adictivo. En tales casos, quizá sea difícil saber si la inversión de preferencias se debe a la estructura de descuento o a otros factores. Una decisión de ayunar que se toma con el estómago lleno puede disiparse tan pronto como la persona comienza a sentir hambre otra vez. Una decisión de dejar de fumar puede debilitarse ante la visión de otra persona que enciende un cigarrillo. Este fenómeno se denomina *dependencia de señales*: anhelos desencadenados por señales visuales asociadas al consumo de la sustancia adictiva. Una decisión de dejar de jugar por dinero que es provocada por los sentimientos de culpa del apostador con respecto a su familia puede perder firmeza una vez que la fuerza de la emoción se desvanece (Capítulo 8). Quizá sea difícil, también, decir si estamos ante una dilación o ante factores viscerales. La decisión de

tomar una medicación regular puede perder vigor a causa del declive de las emociones fuertes que en un primer momento impulsaron al paciente a consultar a un médico.

Una vez que el agente entiende que está sujeto a esos otros mecanismos, quizás se precomprometa a obstaculizar su funcionamiento. Para impedir que el hambre debilite su determinación de hacer dieta, tal vez tome una píldora que atenúa el ansia de comida. De manera más drástica, el agente puede hacer que le fijen la mandíbula, de manera que sólo esté en condiciones de ingerir líquidos. Si sabe que su deseo de postres depende de la existencia de señales, no irá a restaurantes donde aquellos se exhiban en mesas rodantes.[71] Los ex adictos a la heroína se mantendrán alejados de los lugares donde consumían la droga. Los ex jugadores aprenden que no deben ir al casino «sólo para ver jugar a los otros». Si el agente puede predecir que su ira se desvanecerá, de modo que ya no quiera castigar al ofensor, tal vez decida aplicar el castigo de inmediato. Este comportamiento se observó en Bélgica luego de la guerra. Sobre la base de su experiencia en la Primera Guerra Mundial, muchos belgas creían que al cabo de un tiempo la voluntad popular de aplicar severas sentencias a los colaboracionistas cedería su lugar a la indiferencia. En consecuencia, querían que los juicios se sustanciaran lo más rápidamente posible, antes de que las emociones se disiparan.

En la lucha contra la adicción, es muy común la estrategia de imponerse costes. Cuando el general De Gaulle quiso dejar de fumar, contó su intención a todos sus amigos, para incrementar los costes de la reincidencia. En un centro de tratamiento de adictos a la cocaína, en Denver, se brinda a los médicos la oportunidad de escribir una carta autoacusatoria a la Junta Estadual de Examinadores Médicos, en la que confiesan su consumo de drogas y solicitan que les retiren la licencia. La carta se despacha automáticamente si los análisis revelan que han tomado cocaína. Algunos ex alcohólicos procuran mantenerse sobrios tomando disulfiram, una droga cuyo efecto consiste en suscitar un violento malestar en el consumidor si bebe un trago.[72]

Las postergaciones autoimpuestas también pueden ser efectivas para resistir los anhelos vehementes. Para impedirme beber impulsivamente, puedo guardar mis botellas de licor en una caja fuerte con un dispositivo cronometrado de apertura. O bien puedo adoptar la política de no tener licor en casa, de modo que deba ir a una tienda a comprarlo. De hecho, la técnica

del disulfiram combina la imposición de costes con postergaciones, dado que una vez que se toma la píldora hay que dejar pasar dos días sin volver a tomarla, antes de poder beber sin sentir náuseas. El centro de adicción a la cocaína también combinaba costes con postergaciones. Permitía a la gente liberarse del convenio mediante la firma de una declaración certificada de retiro del plan. Se establecía un plazo de dos semanas. Quien presentara una solicitud de retiro podía recuperar la carta autoincriminatoria al cabo de ese lapso. Pero si durante este período el retiro se cancelaba, para volver a pedirlo hacía falta otra notificación con dos semanas de anticipación. Aunque muchos de los pacientes apelaron al procedimiento de retiro, ninguno pasó dos semanas sin revocar la revocación.

El interés por el precompromiso, en contraste con la inconsistencia temporal y el comportamiento excesivo, es relativamente reciente. Los autores clásicos que se refirieron al tópico se concentraban en el compromiso previo contra la *pasión*, tomada en un sentido lato que también incluye la intoxicación y los estados psicóticos.[73] En la *Odisea*, Homero presentó un caso que se ha convertido en el ejemplo típico del precompromiso: Ulises al atarse al mástil para ser incapaz de responder al canto de las sirenas. En *De la cólera*, Séneca escribió:

> Mientras estemos en nuestro sano juicio, mientras seamos nosotros mismos, pidamos ayuda contra un mal que es poderoso y al que a menudo nos entregamos. Quienes son incapaces de beber con discreción y temen que la copa los haga atolondrados e insolentes, instruyan a sus amigos para que los aparten del festín; quienes han advertido que la náusea los hace irrazonables, den órdenes de que en tiempos de enfermedad no se los obedezca.

En *La Princesse de Clèves*, la novela de Madame de Lafayette, la princesa escapa de la corte para evitar la tentación de responder a las insinuaciones del duque de Nemours; y aun más adelante, cuando su esposo ya ha muerto y ella es libre de volver a casarse, se mantiene al margen. «Sabedora de que las circunstancias afectan las más prudentes resoluciones, no estaba dispuesta a correr el riesgo de tener que modificar la suya o de regresar al lugar donde vivía el hombre a quien había amado.» En *Lucien Leuwen*, la novela de Stendhal, Madame de Chasteller toma la precaución de ver a Lucien únicamente en presencia de una dama de compañía, con el fin de que su abandono al amor que siente por él le resulte de un coste prohibitivo.

Estas estrategias son bastante habituales. Cuando la gente quema sus puentes, tal vez lo haga por razones estratégicas (Capítulo 19), pero en

ocasiones lo hace para no caer en las garras del miedo. Quizá me mantenga al margen de la fiesta de mis compañeros de oficina porque sé, debido a pasadas experiencias, que probablemente tome algunos tragos de más y que, a causa de su efecto desinhibidor, el alcohol me impulsará a exhibir un comportamiento agresivo o seductor que más adelante lamentaré. O bien puedo decidir llevar a mi esposa, para elevar el coste de un comportamiento semejante. Es menos probable que la mera *resolución* de no beber (o de no dejarme llevar por la emoción si lo hago) sea eficaz, visto «el poder de la situación» (Capítulo 10). De manera similar, el control de la ira mediante el dispositivo intrapsíquico de contar hasta diez antes de replicar o atacar ferozmente presupone un desapego que suele faltar en el ardor del momento. Una recomendación general de autodefensa significa de hecho «romper la cadena con anticipación» en vez de contar con el autocontrol frente a la tentación o la provocación. Como dijo Mark Twain: «Es más fácil quedarse afuera que salir». Para casos extremos, remito a un titular de *The New York Times* del 5 de abril de 1996: «Texas accede a cirugía para un abusador: próximo a salir de la cárcel, el hombre solicita la castración para doblegar sus urgencias sexuales».

Las estrategias de postergación podrían parecer las más prometedoras para enfrentarse a la irracionalidad basada en las emociones. Como éstas suelen tener una vida media breve, cualquier obstáculo a la ejecución inmediata de una tendencia a la acción podría ser un remedio eficaz. Como he señalado con anterioridad, las autoridades públicas cuentan de hecho con esta característica de la emoción cuando piden a la gente un tiempo de espera antes de tomar ciertas decisiones importantes. Es poco frecuente, sin embargo, que la gente se imponga *a sí misma* postergaciones con el objeto de contrarrestar la pasión. Es posible que las tecnologías necesarias sencillamente falten. Un ejemplo, con todo, es el «matrimonio pactado» que ofrecen tres estados de los Estados Unidos (Arkansas, Arizona y Luisiana), una forma opcional de matrimonio al que es más difícil entrar y del que es más difícil salir que del matrimonio convencional. Por lo común, una pareja que ha hecho un matrimonio pactado sólo puede divorciarse al cabo de dos años de separación, en contraste con el tiempo de espera normal de seis meses. La pequeña minoría (menos del 1% de las parejas que se casan) que utiliza esta opción lo hace, presuntamente, para señalar su compromiso recíproco y para protegerse de las pasiones y tentaciones efímeras.

Con frecuencia, el precompromiso implica la ayuda de otros individuos, organizaciones o autoridades públicas. Es necesario, sin embargo, que éstas sean independientes del agente que emite las instrucciones del precompromiso, pues de lo contrario él podría desautorizarlas. Con el fin de luchar contra su adicción al opio, Samuel Coleridge contrató a un hombre para que le impidiera por la fuerza la entrada a las boticas. Una vez que el hombre intentó frenarlo, Coleridge dijo:

> Oh, qué desatino. Ha surgido una emergencia, una espantosa emergencia, del todo inesperada. No importa lo que le haya dicho en épocas ya remotas. Lo que le digo *ahora* es que, si no aparta los brazos de la entrada de este muy respetable boticario, tendré buenos fundamentos para accionar contra usted por asalto y agresión.

De manera similar, Mao Tse-tung dio la orden de que se ignorara cualquier orden que impartiera después de haber tomado somníferos. Cuando, luego de tomar una de esas píldoras, ordenó a su asistente que enviara una invitación al equipo estadounidense de tenis de mesa para visitar China (ése fue el comienzo de las relaciones chino-norteamericanas), y el asistente le preguntó: «¿Debo tomar en serio sus palabras?», Mao le contestó: «Sí. Hazlo pronto, de lo contrario no habrá tiempo».

Las organizaciones son herramientas más confiables para autoobligarse. La clínica de Denver para adictos a la cocaína y los clubes de Navidad ofrecen opciones de autoobligación que los individuos no podrían encontrar por su cuenta y que están deliberadamente destinadas a ayudarlos a superar sus problemas[74] *e* impedirles rescindir sus instrucciones. En Noruega, la ley de protección de la salud psíquica autoriza a los individuos a comprometerse *en forma voluntaria pero irreversible* a someterse a un tratamiento de tres semanas en una institución psiquiátrica. Al parecer, sin embargo, el sistema no funciona, porque los médicos tienen el derecho pero no la obligación de retener a los individuos una vez que éstos están en la clínica. Para que la medida fuera eficaz, debería permitirse a los pacientes demandar a su hospital si, a su pedido, éste les diera un alta prematura.

En 1996, el estado de Misuri puso en marcha un programa de autoexclusión destinado a los jugadores compulsivos. Quien se inscribe en una lista de autoexclusión tiene prohibida de por vida la entrada a los casinos flotantes del estado. Si el individuo, pese a todo, trata de ignorar la prohibición y apuesta en cualquiera de esos establecimientos, debe ser expulsado del barco y «el licenciatario colaborará con el agente de la comisión para denunciar el incidente a la autoridad procesal

correspondiente y solicitar la presentación de una demanda […] por invasión de propiedad privada, un delito menor». Se debe negar toda ganancia a los jugadores autoexcluidos, si se las han ingeniado de algún modo para subir a bordo, jugar y ganar. El Estado también puede asumir un papel más activo e imponer plazos a los abortos, las compras de armas o los divorcios (¡y los matrimonios!) y otorgar a los consumidores un período de espera de tres días o una semana, durante el cual éstos pueden cancelar las compras hechas en un momento de entusiasmo. Sin embargo, éstos son ejemplos de *paternalismo estatal* y no de *autopaternalismo asistido por el Estado*.

A veces, las constituciones políticas se conciben como dispositivos de precompromiso o una forma de *autopaternalismo colectivo*. En 1871, John Potter Stockton escribió que «las constituciones son cadenas con las que los hombres se amarran en los momentos de cordura para no morir por mano propia en el día de su frenesí». Otra metáfora habitual es la que sostiene que las constituciones son lazos impuestos por Pedro, cuando está sobrio, a Pedro, cuando esté ebrio. El bicameralismo suele mencionarse como un ejemplo de compromiso político anticipado: al hacer que todas las leyes pasen por dos cámaras, se da tiempo para que las pasiones impulsivas se enfríen y la razón (¡o el interés!) recobre su ascendencia. La imposición de plazos para la vigencia de las enmiendas constitucionales se ha justificado con el mismo argumento. No obstante, si el precompromiso se entiende como una *auto*obligación, la extensión del caso individual al caso colectivo y del caso intrageneracional al caso intergeneracional es bastante dudosa. No solemos ver comunidades que se obligan a sí mismas, sino mayorías que obligan a minorías y generaciones presentes que obligan a generaciones futuras. Por otra parte, como las constituciones se redactan habitualmente en tiempos turbulentos, sus mismos redactores o fundadores están a menudo en las garras de la pasión. Al estar «ebrios», quizá no adviertan la necesidad de tomar precauciones contra la ebriedad. Así, el 7 de septiembre de 1789, cuando la Asamblea Constituyente francesa debatía la incorporación del unicameralismo o el bicameralismo a la constitución, el diputado Adrien Duquesnoy escribió la siguiente entrada en su diario:

> Si es lícito hacer una evaluación de probabilidad, parece evidente que la mayoría de la asamblea nunca votará a favor de las dos cámaras. Este resultado quizá tenga grandes desventajas, pero la situación es tal y los espíritus están tan exaltados, que ningún otro es posible; tal vez sea posible hacer un cambio dentro de algunos años. Llegaremos a entender entonces que una asamblea

única, en una nación tan extremadamente impetuosa como la nuestra, puede producir los más terribles efectos.

* * *

NOTA BIBLIOGRÁFICA

El dispositivo intrapsíquico del agrupamiento o la acumulación de las opciones ha sido extensamente analizado por George Ainslie, *Picoeconomics: The Strategic Interaction of Sucessive Motivational States within the Person*, Cambridge y Nueva York, Cambridge University Press, 1992. Se encontrará un intento de dar fundamentos racionales a esas «reglas privadas» en Roland Bénabou y Jean Tirole, «Willpower and personal rules», *Journal of Political Economy*, 112(4), 2004, págs. 848-886. La discusión sobre la miopía decisional se basa en Ole-Jørgen Skog, «Hyperbolic discounting, willpower, and addiction», en Jon Elster (comp.), *Addiction: Entries and Exits*, Nueva York, Russell Sage, 1999, págs. 151-168. Las respuestas estratégicas dadas por individuos sofisticados que son conscientes de su propensión a descontar hiperbólicamente el futuro se examinan en Ted O'Donoghue y Matthew Rabin, «Doing it now or later», *American Economic Review*, 89(1), 1999, págs. 103-124. La idea del precompromiso o la autoobligación para hacer frente a las inclinaciones irracionales propias se analiza en Thomas Schelling, «Egonomics, or the art of self-management», *American Economic Review: Papers and Proceedings*, 68(2), 1978, págs. 290-294, y en varias de las publicaciones posteriores del mismo autor. Yo la he discutido en *Ulysses and the Sirens: Studies in Rationality and Irrationality*, edición revisada, Cambridge y Nueva York, Cambridge University Press, 1984 [*Ulises y las sirenas: estudios sobre racionalidad e irracionalidad*, México, Fondo de Cultura Económica, 1989], en *Ulysses Unbound: Studies in Rationality, Precommitment and Constraints*, Cambridge y Nueva York, Cambridge University Press, 2000 [*Ulises desatado: estudios sobre racionalidad, precompromiso y restricciones*, Barcelona, Gedisa, 2002], y en «Don't burn the bridge before you come to it: ambiguities and complexities of precommitment», *Texas Law Review*, 81, 2003, págs. 1751-1788 [«No queme sus puentes antes de llegar a ellos: algunas ambigüedades y complejidades de los compromisos previos», en Antanas Mockus, Jon Elster y otros, *Acción colectiva, racionalidad y compromisos previos*,

Bogotá, Universidad Nacional de Colombia, 2006]. Un libro dedicado al tratamiento de la omisión de tomar medicamentos recetados es el de Gérard Reach, *Pourquoi se soigne-t-on: une esquisse philosophique de l'observance*, París, Éditions Le Bord de l'Eau, 2005. La historia sobre Coleridge se encontrará en Thomas de Quincey, *Confessions of an Opium Eater*, Londres, Penguin Books, 1968, pág. 145 [*Confesiones de un inglés comedor de opio*, Madrid, Cátedra, 1997]. La historia sobre Mao se ha tomado de Jung Chang y Jon Halliday, *Mao: The Unknown Story*, Nueva York, Alfred Knopf, 2005, págs. 580-581 [*Mao: la historia desconocida*, Madrid, Taurus, 2005]. Si bien en *Ulises y las Sirenas* me mostré entusiasta con respecto a la idea de las constituciones como dispositivos de precompromiso, en *Ulises desatado* me retracté.

Notas al pie

[63] Los estudiosos de la economía, en particular, parecen comportarse de este modo.

[64] Para verlo ilustrado, supongamos que el valor presente de una unidad de utilidad en el tiempo futuro *t* es 1/(1 + *t*) y que el agente se expone dos veces a la elección entre una recompensa pequeña de 3 y una grande de 10. Las pequeñas recompensas están disponibles en los tiempos 0 y 6, y las grandes, en los tiempos 3 y 9. En el tiempo 0, el valor presente de la primera recompensa grande es 10/(1 + 3) = 2,5, que es menos que el valor presente (instantáneo) de la recompensa pequeña. Si la elección se hace sobre la base exclusiva de esta comparación, se escogerá la recompensa pequeña. La misma elección se hará, por idénticas razones, en el tiempo 6. Como la suma de los valores presentes de las dos recompensas pequeñas es 3 + 3/(1 + 6) 3,43 □□, y la suma de los valores presentes de las dos recompensas grandes es 10/(1 + 3) + + 10/(1 + 9) = 3,5, el agrupamiento llevará al agente a preferir las dos recompensas grandes.

[65] Si el agente padece de «miopía decisional» (Capítulo 12), el agrupamiento tal vez no funcione. Supongamos que el agente agrupa sus elecciones bastante antes de que la primera recompensa pequeña esté disponible. En el momento del agrupamiento, el valor presente del flujo de recompensas grandes es mayor que el del flujo de recompensas pequeñas, y el agente tiene el firme propósito de esperar la primera recompensa grande. Al acercarse el momento de tener a su alcance la primera recompensa, esa intención puede o no sobrevivir. Esta inversión de preferencias no se debe al descuento hiperbólico de por sí, sino a la miopía decisional.

[66] Una analogía: supongamos que, al votar, yo puedo inducir a muchos otros, que de lo contrario se abstendrían, a hacer lo mismo. Adviértase que según este supuesto, no interviene ningún pensamiento mágico: sólo hay un efecto multiplicador causal.

[67] En la votación, el efecto del pensamiento mágico consiste en ayudarnos a superar la propensión a adoptar un comportamiento racional socialmente nocivo, y no a contrarrestar la irracionalidad.

[68] Tal vez crean, erróneamente (Capítulo 10), que la consistencia a través de las situaciones provocará un desencadenamiento transituacional de fracasos.

[69] Una ilustración: supongamos que el valor presente de una unidad de utilidad en el tiempo futuro *t* es 1/(1 + *t*), y que sufriré cada vez más si postergo mi visita al dentista: si voy hoy, sufriré un dolor de – 2,75; mañana será de – 5, y pasado mañana llegará a – 9. Desde la perspectiva de hoy, los valores presentes son, respectivamente, – 2,75, – 5/(1 + 1) = – 2,5, y – 9/(1 + 2) = – 3. Por lo tanto, parecería que la elección óptima desde la perspectiva de hoy es postergar la visita hasta mañana. Sin embargo, como soy sofisticado, sé (hoy) que mañana el valor presente de ir en ese momento será de – 5 y el de ir al día siguiente, de – 9/(1 + 1) = – 4,5, lo que induce una preferencia a esperar hasta ese día. Empero, como hoy prefiero ir hoy a ir pasado mañana, voy hoy.

[70] Una ilustración: supongamos que el valor presente de una unidad de utilidad en el tiempo futuro *t* es 1/(1 + *t*), y que puedo beneficiarme cada vez más si postergo mi consumo de una botella de vino: si la bebo este año obtengo un placer de 2,75; el año que viene será de 5, y dentro de dos años, de 9. Desde la perspectiva de hoy, los valores presentes son, respectivamente, 2,75, 5/(1 + 1) = 2,5, y 9/(1 + 2) = 3. En consecuencia, podría parecer que la elección óptima desde la perspectiva del primer año es beber el vino dentro de dos años. Sin embargo, como soy sofisticado, sé (este año) que el año que viene el valor presente de beberlo en ese momento será de 5, y el de beberlo un año después, de 9/(1 + 1) = 4,5, lo cual genera una preferencia a beberlo el año próximo. Empero, como en el primer año prefiero beberlo el primer año a beberlo el año próximo, lo bebo este año.

[71] Si comparamos este caso con el caso en que el postre tenía que pedirse al comienzo de la comida, podemos ver que ambos implican una autoprotección contra los efectos de la *proximidad de la tentación*, sea espacial o temporal.

[72] La forma más común es la ingesta oral, que hace que el consumo de alcohol provoque náuseas. El implante bajo la piel (ineficaz desde el punto de vista físico, pero psicológicamente eficaz) es menos habitual.

[73] El fenómeno –brevemente mencionado en el texto– consistente en precomprometerse mientras se está en las garras de la pasión y se teme que ésta termine es mucho menos habitual.

[74] En contraste, las cajas fuertes con dispositivos cronometrados de apertura no fueron pensadas para ayudar a la gente a luchar contra sus problemas de bebida.

Capítulo 14

Algunas implicaciones para la interpretación textual

Según una concepción común, la empresa científica tiene tres partes o ramas distintas: las humanidades, las ciencias sociales y las ciencias naturales. A algunos efectos, ésta es una manera provechosa de dividir el campo de la ciencia, pero a otros efectos una distinción rígida puede impedir la fertilización cruzada. En la cuarta parte sostendré que las ciencias sociales pueden beneficiarse con el estudio biológico de los seres humanos y otros animales. En este capítulo, argumento que las humanidades y las ciencias sociales tienen más en común de lo que suele suponerse. En particular, procuraré mostrar que la *interpretación* de las obras de arte y la *explicación* son empresas que guardan estrecha relación. Entender una obra de arte es explicarla en términos de los estados mentales antecedentes de su creador. Es *lograda* la obra de arte para la cual puede proponerse una explicación basada en la elección racional. Al mismo tiempo, me pronunciaré contra lo que podríamos llamar «interpretación por las consecuencias», una frase que se aclarará más adelante. La descripción que haré no abarca todas las formas artísticas. Aun dentro de la literatura, ámbito al cual me limitaré, sólo tiene sentido para las novelas y obras clásicas (anteriores a 1850), definidas por la convención tácita de que los sucesos y personajes que se describen *podrían haber sido reales*.[75]

Consideremos en primer lugar la racionalidad como un motivo de los *personajes* de la ficción o las piezas teatrales. Un problema clásico de la crítica literaria es por qué Hamlet se demora en vengar la muerte de su

padre. Se han propuesto muchas explicaciones. Algunas de ellas apelan a la irracionalidad y aluden a la debilidad de la voluntad o la depresión clínica. También hay, empero, una descripción simple basada en la elección racional. Aunque en un comienzo Hamlet cree lo que el espectro de su padre le dice acerca de Claudio, luego decide *recoger más información* mediante la puesta en escena de una obra que será «el lazo en que se enrede la conciencia del rey». Sin embargo, una vez que las reacciones de éste confirman su creencia, carece de una *oportunidad* para realizar su deseo, consistente en que Claudio arda eternamente en el infierno. Aunque ha tenido una ocasión de matarlo en el momento de la oración, de haberlo hecho habría asegurado, de acuerdo con la teología de la época, la salvación de su víctima y no su condena. Luego mata a Polonio detrás de un cortinado, creyendo *errónea pero no irracionalmente* que es el rey. Dada la información con que contaba, su creencia de que era el rey quien estaba escondido detrás del cortinado era racional. Por otra parte, no tenía motivos para recoger *más* información, dado que podía suponer razonablemente que quien se ocultara de ese modo en presencia de la reina debía ser el rey.

No pretendo que ésta sea la interpretación correcta (de hecho, todavía no he dicho qué significa que una interpretación sea «correcta»). Mi observación es simplemente que los tres episodios que he mencionado son prima facie coherentes con la idea de que Hamlet persigue en forma racional la meta de vengar la muerte de su padre. Otra cuestión es si la idea es coherente con las reiteradas acusaciones que Hamlet se hace a sí mismo, en el sentido de carecer de la resolución necesaria para tomar venganza. Muchos comentaristas interpretan esos famosos monólogos como un signo de debilidad de la voluntad y estiman que los primeros dos episodios se basan en excusas autoengañosas para la inacción. (El tercer episodio es más difícil de acomodar a ese punto de vista.) Ahora bien, aunque la debilidad de la voluntad y el autoengaño transgredan los cánones de la racionalidad, son perfectamente *inteligibles* (Capítulo 3). Y cuando nos ocupamos del desarrollo interno de la obra de arte, la idea más útil para la tarea interpretativa es la inteligibilidad y no la racionalidad.

En contraste con el punto de vista interno, podemos adoptar el punto de vista externo del autor. A la pregunta: «¿Por qué Hamlet demora su venganza hasta el quinto acto?», podríamos responder: «La muerte del rey debe producirse al final de la obra».[76] Esta cuestión tiene que ver con la construcción dramatúrgica, no con la psicología. En sí misma, aquella

respuesta no sería satisfactoria. Si Shakespeare hubiera demorado la venganza por medio de una serie de sucesos arbitrarios o coincidencias ad hoc, con el mero objeto de hacer que se produjera al final de la obra, lo habríamos visto como un fracaso autoral. Más directamente, habría sido un caso de *irracionalidad autoral*.

La racionalidad autoral es como la atribuida a Dios. Al igual que Dios, el autor pone en marcha un proceso en el cual cada suceso puede *explicarse dos veces*, primero causalmente y luego teleológicamente. Tomo esta idea de Leibniz, quien escribió que hay

> dos reinos, uno de causas eficientes, otro de causas finales, cada uno de los cuales, por separado, basta en detalle para dar una razón del todo, como si el otro no existiera. Pero ninguno de los dos es adecuado sin el otro cuando consideramos su origen, pues emanan de una fuente en la que el poder que hace las causas eficientes y la sabiduría que rige las causas finales se encuentran unidos.

La meta de Dios es crear el mejor de todos los mundos posibles. Si se la precisa para incluir la dimensión temporal, la idea puede entenderse como *la mejor de todas las secuencias posibles*. Si bien la transición de un estado del universo al siguiente se produce en virtud de la causalidad física ordinaria, el estado inicial y las leyes de causalidad se han elegido para maximizar la perfección global de la secuencia.

Si nos limitamos al teatro o la novela clásicos, la tarea del autor consiste en desarrollar la trama por medio de lo que los personajes dicen y hacen, a menudo en respuesta mutua. El objetivo es hacerlo de tal forma que el valor estético se maximice. Así, cada acción o enunciado de un personaje puede explicarse dos veces, como una reacción a acciones y enunciados previos (o sucesos externos) y como un generador de sorpresa, tensión y, en última instancia, resolución de ésta en el lector. La primera explicación se apoya en la inteligibilidad de los personajes, la segunda, en la racionalidad del autor, en un sentido que ahora trataré de aclarar.

El hecho de que los autores hagan con frecuencia muchos borradores antes de sentirse satisfechos o de dejar la pluma, es una prueba irrefutable de que están embarcados en un proceso de *elección* y de que tienen criterios explícitos o implícitos de *perfeccionamiento*. Las *pequeñas* variaciones que esos borradores suelen contener sugieren que los autores aspiran a un *máximo local* de la forma de perfeccionamiento por el cual se afanan. Sin embargo, la diferencia entre un autor y alguien que no hace sino subir a lo largo de un gradiente es que la *creatividad* del primero va más allá de la

mera elección.[77] La razón por la cual la creación de una obra literaria no puede reducirse a la elección racional es que el número de secuencias significativas de palabras es demasiado grande para que una persona las explore todas y seleccione «las mejores». Si bien un «creador racional» puede procurar que un problema sea más fácil de abordar excluyendo deliberadamente algunas secuencias (ésa es una de las funciones del metro y la rima en el verso), por lo común quedarán demasiadas opciones para que la elección sea viable como mecanismo de selección. El autor deberá apoyarse, en cambio, en su maquinaria asociativa inconsciente.

En consecuencia, la creación racional tiene que ver en gran medida con acertar el segundo decimal o, para cambiar la metáfora, con trepar hasta la cumbre de la colina más cercana. Para decirlo con una metáfora más, ésta es una tarea del hemisferio izquierdo. La tarea de acertar el primer decimal o encontrar una colina que descuelle sobre las demás, que corresponde al hemisferio derecho, no está dentro del alcance de la racionalidad. No obstante, aun reducida a la tarea de la sintonía fina, la racionalidad autoral importa. Tal cual lo sugiere la expresión «obra maestra menor», tal vez sea mejor alcanzar la cumbre de una colina baja que quedarse en las laderas de una más alta. Sin dar a entender ningún juicio comparativo, *Crónica de una muerte anunciada* y *El ángel que nos mira* pueden servir para ilustrar las dos posibilidades.

Querría enumerar y luego examinar algunas exigencias que la racionalidad impone al autor. Primero, los actos y los dichos de los personajes tienen que ser inteligibles. Segundo, el autor tiene que satisfacer los requisitos gemelos de la *plenitud* y la *parsimonia*. Tercero, la obra tiene que transcurrir con *fluidez*, en el sentido de minimizar el recurso a accidentes y coincidencias. Cuarto, tiene que proponer un patrón psicológicamente gratificante de la acumulación y resolución de la tensión.

La inteligibilidad puede ser absoluta o relativa, y, en este último caso, global o local. El interrogante de la inteligibilidad absoluta es si *algún* ser humano puede comportarse de ese modo. El interrogante de la inteligibilidad relativa global es si el comportamiento de una persona de ficción es coherente con su carácter global tal como lo ha exhibido en momentos anteriores de la obra. El interrogante de la inteligibilidad relativa local es si el comportamiento de una persona de ficción es coherente con su conducta en situaciones anteriores de la obra. Mientras que los requisitos de inteligibilidad absoluta e inteligibilidad relativa local son limitaciones

cruciales impuestas a la racionalidad autoral, el de la inteligibilidad relativa global no lo es. Si se quiere, el respeto por la última limitación puede verse como un defecto estético.

En algunos casos, puede transgredirse la inteligibilidad absoluta por exceso de racionalidad. Consideremos otra vez a la Medea de Eurípides o la Fedra de Racine, ambas igualmente lúcidas en lo concerniente a sus pasiones autodestructivas. Se las pinta como sometidas a la debilidad de la voluntad en sentido estricto, sabedoras de que lo que hacen es contrario al juicio global que sostienen *en el momento mismo de actuar*. Aunque la pasión las lleva a desviarse de ese juicio, no lo afecta. La Hermione de Racine es un personaje más creíble. Como las emociones le nublan el juicio, es más ilusa que débil de voluntad. Mi sugerencia (no es más que eso) es que la presencia simultánea de una emoción extrema y una plena lucidez cognitiva contraría lo que sabemos de la naturaleza humana.

Así como demasiada racionalidad puede ser ininteligible, la irracionalidad puede ser perfectamente inteligible. ¿Qué puede serlo más que la reacción del señor de Rênal en *Rojo y negro*, de Stendhal, cuando, ante la presencia de fuertes indicios de que su mujer tiene una aventura con Julien Sorel, elige creer en su fidelidad? El anhelo es el padre del pensamiento. Más paradójicos son los casos en los que el deseo de que nuestra esposa sea fiel genera la creencia de que *no* lo es, a pesar de las pruebas que indican lo contrario. En *Otelo*, «bagatelas tan leves como el aire son para el celoso pruebas tan irrefutables como las afirmaciones de las Sagradas Escrituras». El primero es un caso de cortocircuito, el segundo, un caso de cruce de cables (Capítulo 3).

La inteligibilidad relativa, que en una obra o una novela es transgredida cuando un personaje actúa de manera «poco típica» de sí mismo, plantea problemas diferentes. Primero, debemos tomar en cuenta los argumentos de los psicólogos cuando señalan que los rasgos de carácter tienden a ser *locales* y no globales (Capítulo 10). Si bien muchos escritores (Hamsun menciona a Zola) adhieren a la psicología popular que supone una coherencia a través de las situaciones, los buenos autores (y Hamsun menciona a Dostoievski) no lo hacen. Estos últimos acaso decepcionen a los lectores que esperan que los personajes se comporten «de manera típica», pero dichos lectores no son el público previsto por la obra. Como veremos dentro de poco, aun los buenos autores pueden estar limitados por la psicología defectuosa de sus lectores, pero la creencia en los rasgos

globales no debe hacerse acreedora a su respeto. Sin embargo, los lectores tienen derecho a esperar coherencia local. Si el autor se mete en camisa de once varas, de modo que la única forma de desarrollar la trama tal como se la ha planeado consiste en permitir a un personaje actuar de manera localmente incoherente, viola su contrato implícito con los lectores. Una trama debe desarrollarse a la manera del agua que busca su curso natural cuesta abajo, y no con la intervención del autor que la fuerza a correr cuesta arriba.

Permítaseme ilustrar esta idea con algunos de los comentarios marginales de Stendhal en el manuscrito de *Lucien Leuwen*, su novela póstuma e inconclusa. El escritor hace que el héroe epónimo se enamore de una joven viuda, Madame de Chasteller. Sus sentimientos son correspondidos, pero Lucien no se atreve a acercarse a ella. La misma delicadeza espiritual que lo hace superior al «más consumado Don Juan» y por lo tanto capaz de inspirar amor, también lo hace inferior a cualquier «joven parisino menos educado» que sepa al instante cómo manejar la situación. Para que la trama avance, Stendhal necesita reunir a Lucien y su amada, pero no sabe muy bien cómo hacerlo. Anota en el margen: «Ante lo cual el cronista dice: uno no puede esperar que una mujer virtuosa se entregue por completo; hay que tomarla. El mejor perro de caza no puede hacer más que poner la presa a tiro de fusil. Si el cazador no dispara, el perro no sirve de nada. El novelista es como el perro de su héroe». El comentario ilustra de manera llamativa la necesidad de que el comportamiento de los personajes [*characters*] de una novela sea el «típico» [*in character*] de ellos.

A la larga, Stendhal se las ingenia para maquinar una situación en la cual Lucien y Madame de Chasteller pueden mostrarse y entender su amor recíproco, aunque sin declararlo. Pero sus dificultades no terminan ahí. Su plan para la novela seguía la receta dialéctica de Hollywood: el chico conoce a la chica, el chico y la chica se separan, el chico y la chica vuelven a estar juntos. Como acabamos de ver, Stendhal tropieza con problemas para establecer la tesis. Para producir la antítesis, utiliza un dispositivo ridículo y manifiestamente teleológico: hace que Lucien crea que Madame de Chasteller, a quien él ha visto a diario y de cerca, ha tenido, de improviso, una criatura. Sin embargo, lo que realmente lo deja perplejo es la síntesis. Aunque no sabemos por qué nunca encontró tiempo para escribir la tercera parte, en la cual los amantes debían reunirse, una conjetura es que

su unión no habría sido verosímil. En la segunda parte de la novela, luego de la ruptura, Lucien se convierte en algo parecido a un calavera cínico, fundamentalmente honesto si se lo juzga por los relajados criterios de la monarquía de julio, pero sin duda muy diferente de la torpe persona delicada de quien Madame de Chasteller se ha enamorado. Es posible que Stendhal haya decidido que, de conseguir el amor de la joven, el transformado Lucien violaría la inteligibilidad relativa.

Aristóteles escribió que «la historia […] debe representar una acción, una totalidad completa, en la que los diversos incidentes guarden una relación tan estrecha que la trasposición o eliminación de alguno de ellos significará la desarticulación y disgregación de dicha totalidad. Pues aquello que por su presencia o ausencia no marca una diferencia perceptible no es una verdadera parte del todo». Podríamos interpretar que este pasaje expresa los ideales estéticos de la *plenitud* y la *parsimonia*. El lector está autorizado a pensar que el autor le ha presentado toda la información que necesita para entender el desarrollo de la trama.[78] A la inversa, tiene derecho a suponer que, si el autor le cuenta que llovía cuando un personaje se marchó de la casa, es porque la premisa de la lluvia será necesaria más adelante, y le es lícito creer que un parlamento atribuido a un personaje aspira a decirnos algo sobre la persona o a funcionar como premisa para la acción de otros personajes.[79]

Antes me referí al carácter de «fluidez» de una buena trama, y utilicé como ejemplo la actuación «de manera típica». En términos más generales, las buenas tramas no deben girar en torno de sucesos, accidentes o coincidencias improbables. En *Middlemarch*, el encuentro entre Raffles y el señor Bulstrode, un elemento crucial en el desarrollo de la historia, es tan artificial que nos aparta de la progresión de la novela, que en otros aspectos es fluida. Los accidentes, por supuesto, pueden tener su lugar en una novela. La muerte accidental de un progenitor puede provocar o dar forma al desenvolvimiento de la intriga, como tal vez lo haga la muerte de ambos padres en el mismo accidente. Pero si la trama requiere que mueran en *dos* accidentes separados, se ejerce una violencia sobre la verosimilitud. La muerte conveniente de un cónyuge que permite al héroe o la heroína casarse con su verdadero amor también es el signo de una censurable pereza autoral.

La psicología de los lectores no está, sin embargo, en sintonía fina con la teoría de las probabilidades. Supongamos que el autor tiene, en una

trama, la alternativa de ir de A a B en dos o en seis pasos. Para precisar un poco las cosas, supongamos que los dos pasos requieren sucesos que ocurrirán con una probabilidad de 0,9 y 0,2, respectivamente, mientras que cada uno de los seis sucesos tendrá una probabilidad de 0,75. Si suponemos que los sucesos de cada secuencia son independientes entre sí, la secuencia de dos pasos tiene mayores probabilidades de ocurrir (0,18 contra 0,178), pero se verá que sólo la de seis pasos tiene la propiedad conveniente de ser «fluida». La verosimilitud global de un escenario depende mucho más de la verosimilitud de sus eslabones más débiles que de la cantidad de eslabones. Creo que el autor debe respetar esta peculiaridad de los lectores, pues ella le impide recurrir a fáciles pero improbables coincidencias.

Aun una corriente que circule con fluidez puede tener muchas vueltas y revueltas antes de encaminarse resueltamente hacia el mar. Si no fuera así, la observación de su curso no significaría una experiencia muy provechosa. El autor está obligado, por consiguiente, a brindar las sorpresas necesarias a lectores y espectadores, y a poner los obstáculos necesarios a los personajes, para mantener vivo el interés del público. El repertorio de estratagemas es enorme, y supera cualquier intento de enumeración e incluso de clasificación. Algunas de ellas están íntimamente relacionadas con el género. Dentro del teatro, la comedia, el drama y la tragedia tienen diferentes medios a su disposición. Mientras que la comedia se apoya a menudo en los *malentendidos* para generar tensiones, el drama y más aún la tragedia pueden apoyarse en la *ignorancia*. Cuando los malentendidos se disipan, el resultado es la felicidad; cuando se desvanece la ignorancia, ocurren desastres. Los novelistas pueden sumar su propia voz a las de los personajes para provocar incertidumbre, siempre que no descaminen deliberadamente a los lectores.

Estoy ahora en condiciones de decir a qué me refiero cuando hablo de «interpretación correcta» de un texto. Según lo expuse al comienzo, se trata de una cuestión de explicación. Como todas las explicaciones son causales (incluidas las que mencionan intenciones como causas), y como una causa debe preceder a su efecto, se deduce que las percepciones *reales* del público sobre la obra son estrictamente irrelevantes. Las percepciones previstas, por el contrario, pueden ser parte de la explicación. Entre las causas antecedentes de la obra, la intención autoral no es lo único que importa. Las actitudes inconscientes del autor también pueden influir en ella. Así, es posible que Jules Verne haya dado forma a *La isla misteriosa* no sólo con

sus intenciones antirracistas, sino también con sus prejuicios racistas. Sin embargo, por razón de la brevedad me limitaré a las intenciones conscientes.

La interpretación de una obra literaria es, entonces, la pretensión de que rasgos importantes de ésta pueden atribuirse a decisiones tomadas por el autor con el objeto de realzar el valor estético de la experiencia que un público específico podría esperar obtener de la frecuentación de la obra. Para plantear una idea de este tipo, los críticos literarios deben proceder como lo hacen otros estudiosos. Pueden recurrir a borradores, cuando existen, y a declaraciones del autor sobre la obra, por ejemplo las notas al margen de Stendhal. Pueden recurrir a otras obras del mismo autor, para ver si se observa un patrón similar de elecciones. Pueden remitirse a obras contemporáneas, para distinguir las convenciones que enmarcan las elecciones de las elecciones mismas. Pueden apelar a otras fuentes coetáneas para determinar las expectativas del público que quizá significaron una limitación para el autor.

Al hacer todo esto, su método no difiere en modo alguno del procedimiento utilizado por otros historiadores. Al igual que éstos, los críticos literarios se enfrentan al problema de que los datos son esencialmente finitos, porque el pasado no es susceptible de experimentos. Y al igual que otros historiadores, pueden tratar de minimizar las tentaciones de la «explotación de datos» mediante la triangulación de fuentes antiguas, la búsqueda de nuevas fuentes y el desarrollo de implicaciones novedosas de su interpretación para cotejarlas con las pruebas existentes. La diferencia con otros historiadores puede residir en el hecho de que muchas veces, aunque no de manera invariable, su interpretación va a la par con *juicios de valor*. ¿Tuvo éxito el autor, o estuvo más cerca del éxito que del fracaso, en su objetivo de crear un máximo local de valor estético? Algunos escritores, claro está, no tienen ese objetivo. Tal vez sólo les interese ganar dinero o escribir propaganda, metas que tienen otras exigencias de racionalidad. Pero si se puede plantear un argumento verosímil a favor de la hipótesis de que las aspiraciones del autor eran principalmente de orden estético, es sensato preguntarse, como en el caso de cualquier otro objetivo, en qué medida consiguió realizarlas.

Dije antes que los fracasos autorales pueden ser inteligibles. Los autores, he sostenido, se encuentran bajo una doble presión: necesitan que la trama avance, y deben hacerlo por medio de acciones y enunciados

inteligibles de los personajes. Quizá los censuremos si sacrifican esta última meta a la primera, es decir, si sacrifican la causalidad a la teleología, pero aun así podemos *entender* por qué lo hacen. Aunque no sea verosímil desde el punto de vista causal, la dilación de Hamlet podría parecer teleológicamente inteligible a la luz de la necesidad de Shakespeare de posponer su venganza hasta el final de la pieza. Esto también sería un ejemplo de interpretación. Si bien muy diferente, desde luego, de una interpretación de la dilación en términos de la psicología y las circunstancias de Hamlet, responde a la misma pregunta: ¿por qué esa dilación? A despecho de que en una buena obra literaria todo pueda explicarse dos veces, las obras imperfectas tal vez sólo den cabida a una interpretación.

Querría concluir con un ejemplo de cómo puede la interpretación transgredir o ignorar las exigencias de la explicación. Varios autores recientes han afirmado que Fanny Price, el personaje de *Mansfield Park* de Jane Austen, es maquinadora y estratégica, y que su aparente modestia no es más que una estratagema desplegada para conquistar a Edmund Bertram. También sostienen, por otra parte, que su nombre mismo sugiere «sexo por dinero». *Dos pruebas de intencionalidad invalidan* estas afirmaciones. Primero, en la novela no hay elementos para atribuir intenciones intrigantes a Fanny Price. Si bien su modestia recibe en efecto una recompensa, esa *consecuencia* de su comportamiento no puede explicarla.[80] Segundo, tampoco hay elementos para atribuir a Jane Austen la intención de que sus lectores vieran a Fanny Price como una semiprostituta. Aunque el texto lleve a algunos lectores modernos a hacer esa asociación, los autores en cuestión no aportan pruebas de que Austen pretendiera que su público asociara «Fanny» con la heroína de la novela pornográfica *Fanny Hill* o «Price» [precio] con el sexo pago. Estas «interpretaciones por las consecuencias» tienen mucho en común con las explicaciones funcionales en las ciencias sociales. Se basan en métodos arbitrarios cuyas restricciones no están en los hechos sino únicamente en los límites del ingenio de los estudiosos que las proponen. Sea como fuere, en la cuarta parte examinaremos algunas variedades más respetables de la «explicación por las consecuencias».

* * *

Nota bibliográfica

El enfoque general que adopto en este capítulo es a menudo objeto de la acusación de encarnar una «falacia intencional». Coincido con las respuestas de Noël Carroll a esta crítica, sobre todo las que expone en «Art, intention and conversation», en Gary Iseminger (comp.), *Intention and Interpretation*, Filadelfia, Temple University Press, 1992, págs. 97-131, y en «The intentional fallacy: defending myself», *Journal of Aesthetics and Art Criticism*, 55(3), 1997, págs. 305-309. Dagfinn Føllesdal, «Hermeneutics and the hypothetico-deductive method», en Michael Martin y Lee C. McIntyre (comps.), *Readings in the Philosophy of Social Science*, Cambridge (Massachusetts), The MIT Press, 1994, págs. 233-245, propone una interpretación de *Peer Gynt* basada en un razonamiento similar, con la salvedad de que esta obra no está limitada por la convención de que los sucesos y personajes que se describen puedan haber sido reales. Debo la observación de que la dilación de Hamlet tiene acaso su origen en preocupaciones dramatúrgicas a Edward Wagenknecht, «The perfect revenge –Hamlet's delay: a reconsideration», *College English*, 10(4), enero de 1949, págs. 188-195. Discuto la idea de las obras de arte como máximos locales en el Capítulo 3 de *Ulysses Unbound: Studies in Rationality, Precommitment and Constraints*, Cambridge y Nueva York, Cambridge University Press, 2000 [*Ulises desatado: estudios sobre racionalidad, precompromiso y restricciones*, Barcelona, Gedisa, 2002]. Ese capítulo también contiene un examen más completo de Lucien Leuwen. La idea de las tramas «cuesta abajo» en contraste con las tramas «cuesta arriba» se inspira en Daniel Kahneman y Amos Tversky, «The simulation heuristics», Daniel Kahneman, Paul Slovic y Amos Tversky (comps.), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge y Nueva York, Cambridge University Press, 1982, págs. 201-208. Las interpretaciones de *Mansfield Park* que critico son las de Jill Heydt-Stevenson, «"Slipping into the ha-ha": bawdy humor and body politics in Jane Austen's novels», *Nineteenth-Century Literature*, 55(3), 2000, págs. 309-339, y Jenny Davidson, *Hypocrisy and the Politics of Politeness: Manners and Morals from Locke to Austen*, Cambridge y Nueva York, Cambridge University Press, 2004.

Notas al pie

[75] Uno de los primeros ejemplos de transgresión de esta convención se da hacia el final de *Peer Gynt* (1867), de Ibsen, cuando Peer tiene miedo de ahogarse y el «extraño pasajero» le dice que «uno no muere en medio del quinto acto».

[76] A diferencia de las palabras de Ibsen puestas en labios del «extraño pasajero» (véase la nota anterior), Shakespeare no podría haber hecho decir a Hamlet: «No puedo matar al rey hasta el quinto acto».

[77] Por la misma razón, los intentos de encontrar explicaciones del cambio técnico basadas en la elección racional están condenados al fracaso.

[78] Está claro que algunos detalles potencialmente relevantes pueden excluirse de manera deliberada para dar cierta cabida a la imaginación del lector. La creación racional es compatible con el hecho de que el lector tenga que llenar algunos espacios en blanco (e incluso puede exigir que así sea). Sin embargo, si el artista sobrestima la imaginación de su público, su esfuerzo se juzgará como un fracaso. Supongamos que, para trata de sugerir la incompatibilidad temperamental entre un héroe y una heroína, un novelista hace que los números de las direcciones de sus casas sean primos entre sí, es decir sin divisores comunes. Exceptuando circunstancias especiales, no podrá contar razonablemente con que el lector sea capaz de advertir ese hecho.

[79] La redundancia, desde luego, no siempre debe evitarse, dado que puede cumplir una función estética. Para dar la idea de aburrimiento, aquella puede ser más eficaz que una mera declaración del autor. Aun en ese caso, sin embargo, habría un punto en que la repetición terminaría por aburrir al *lector* en vez de evocar el aburrimiento del personaje.

[80] Además, la hipótesis de una Fanny Price mercenaria no puede explicar su rechazo de una propuesta de matrimonio hecha por Henry Crawford, hombre de mejor posición.

CUARTA PARTE

Lecciones de las ciencias naturales

Las diversas disciplinas científicas pueden mantener entre sí una de dos relaciones: *reducción* o *analogía*. La reducción adopta la forma de la explicación de fenómenos de un nivel de la jerarquía de las ciencias en términos de fenómenos de un nivel inferior (véase la Figura IV.1).

Los programas reduccionistas suelen despertar polémicas. Durante mucho tiempo, numerosos biólogos afirmaron con vehemencia que la reducción de la biología a la química no podía funcionar de ningún modo, pero funcionó. Desde Émile Durkheim en adelante, muchos han sostenido que no hay posibilidades de reducir la ciencia social a la psicología. Una idea central de la presente obra es que sí las hay.

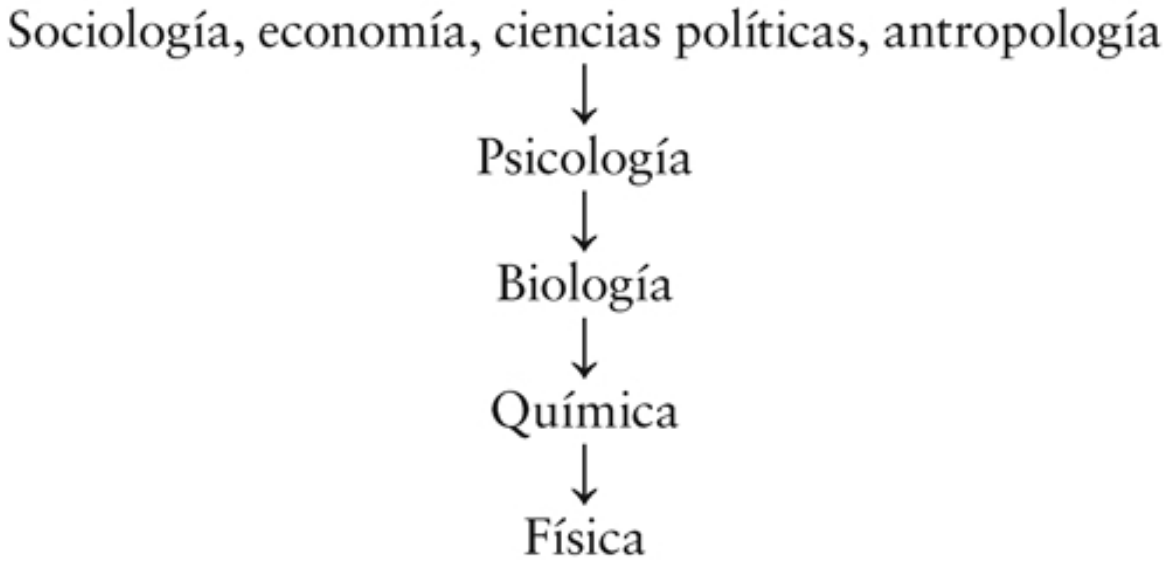


FIGURA IV.1

Entre estas dos reducciones está la reducción de la psicología a la biología. Las disciplinas biológicas relevantes son la genética, la fisiología, la biología del desarrollo y la biología evolutiva. Las dos primeras estudian las causas próximas de la estructura y el comportamiento de los organismos; las dos últimas, las causas remotas en la historia del organismo individual o de la especie. El estudio de la estructura y el estudio del comportamiento están relacionados, en cuanto la estructura proporciona al comportamiento tanto oportunidades como restricciones. El hecho de que tengamos dos riñones y sólo necesitemos uno nos permite donar el otro para trasplantarlo a un hermano y hace posible la aparición de normas sociales

que imponen o proscriben esa práctica. No obstante, la razón por la que tenemos dos riñones no es la de permitir el trasplante de donantes vivos. Muchas estructuras existen por lo que nos permiten hacer, pero éste no es un ejemplo de ello. Con frecuencia, sin embargo, es difícil decir si los efectos facilitadores de una estructura son accidentales o explicativos.

La relevancia de la biología para las ciencias sociales debería ser evidente, dado que sus dominios se superponen. No obstante, muchos especialistas en ciencias sociales se resisten a admitir las explicaciones biológicas con el argumento de que son «reduccionistas». La acusación es extraña para quienes creen, como yo, que el reduccionismo es el motor del progreso en la ciencia. Con todo, si se agrega a «reduccionismo» un modificador como «prematuro», «crudo» o «especulativo», la objeción puede estar bien fundada.

El reduccionismo *prematuro* se observa cuando estudiosos que están convencidos de la factibilidad última del pasaje de las explicaciones de un nivel superior a las explicaciones de un nivel inferior tratan de dar ese paso antes de establecer las técnicas de medición, los conceptos y las teorías necesarias. Un ejemplo clásico es la fisiología mecanicista de Descartes, que Pascal comentó en los siguientes términos: «Descartes. Es preciso decir en grueso: "Esto se hace por figura y movimiento", porque es verdad. Pero decir cuáles y componer la máquina, esto es ridículo; porque esto es inútil, incierto y penoso». En nuestros días, quienes proponen algoritmos para el reconocimiento de patrones y la traducción automática quizá se encuentren en una situación similar. Las tareas de reconocer una cara humana y de detectar frases desatinadas, que nosotros realizamos sin esfuerzo, están hasta ahora más allá de la capacidad de los sistemas artificiales.

El reduccionismo *crudo* se observa cuando los estudiosos tratan de explicar un comportamiento específico en términos biológicos, en vez de explicar la *capacidad* o la *tendencia* correspondientes a ese comportamiento, que en un caso dado pueden o no utilizarse o realizarse. Un ejemplo es el intento de explicar el comportamiento político en función del «imperativo territorial» verificado en animales inferiores. Otro es la idea de que la práctica de la halterofilia puede explicarse como resultado de la selección sexual, análoga a las plumas del pavo real o las cornamentas gigantes en el ciervo. Podrían mencionarse muchos otros casos.

El reduccionismo *especulativo* se observa cuando los estudiosos producen relatos del tipo «es así, porque así es» que describen cómo *podría*

haber aparecido un comportamiento determinado, sin mostrar que apareció *efectivamente* de esa manera. La sociobiología y el campo estrechamente relacionado de la psicología evolutiva abundan en ejemplos, entre ellos la idea de que el autoengaño se ha desarrollado por sus beneficios evolutivos o la de que la depresión posparto en las mujeres evolucionó como una herramienta de negociación (Capítulo 17).

Decir que el mal reduccionismo es malo no es muy esclarecedor. El deseo de reducir fenómenos complejos a fenómenos más simples puede asumir formas simplistas, pero otro tanto puede ocurrir con cualquier estrategia de investigación en la ciencia. La historia de ésta muestra con evidencia abrumadora que el reduccionismo es una fuerza progresiva y que el antirreduccionismo es una fuerza obstruccionista en el campo científico. La historia también muestra el riesgo de utilizar *analogías* entre una disciplina científica y otra para generar hipótesis. En sí mismo, el uso de analogías es inofensivo: las hipótesis científicas deben juzgarse por sus descendientes (implicaciones verificables), no por sus ancestros. No obstante, cuando el pensamiento analógico lleva a los estudiosos a privilegiar un tipo de hipótesis sobre otros, el resultado suele terminar en el gabinete de los horrores del pensamiento científico. La analogía entre la sociedad y los organismos biológicos, por ejemplo, se ha utilizado para respaldar la idea de que las sociedades, como los organismos, son entidades autorreguladoras con mecanismos incorporados de corrección homeostática (a saber, las revoluciones, entre otros). En el siglo XIX, los académicos procuraban determinar en sus debates qué elemento de la sociedad equivalía a la *célula* en el organismo, sin preguntarse si había acaso razones para esperar alguna analogía. En vez de apelar a las analogías orgánicas, otros autores se han valido de las analogías físicas y han buscado el equivalente social de las leyes de Newton o la fuerza de la gravedad. Los investigadores que sostienen que las ciencias sociales pueden ejercer influencia sobre el objeto que estudian invocan de manera habitual el principio de incertidumbre de Heisenberg, como si su profundidad pudiese transformar su perogrullada en algo igualmente profundo.

En el Capítulo 15 considero algunos descubrimientos de la fisiología y la ciencia del cerebro que prometen ofrecer descripciones reduccionistas de fenómenos como el miedo, la confianza y la tendencia a «sacar conclusiones precipitadas». En el Capítulo 16 y en el Capítulo 17 examino la teoría de la selección natural, a la que se ha dado el doble uso de la

reducción *y* la analogía. Sostengo que si bien algunos intentos reduccionistas son plausibles, otros son prematuros, crudos o especulativos. El uso de la selección natural como analogía de los fenómenos sociales tiene un valor más dudoso. Si en el Capítulo 16 expongo los mecanismos de la selección natural de una manera que acaso parezca detallada en exceso para un libro dedicado a la ciencia social (aunque demasiado superficial para servir como exposición del tema), lo hago, entre otras razones, para mostrar que en el mundo social no existe nada que sea siquiera remotamente comparable. Las discusiones acaloradas de los especialistas en ciencias sociales sobre «selección social» y «evolución social» tienen demasiado poca precisión y escasa sustancia para tomarlas en serio.

* * *

NOTA BIBLIOGRÁFICA

Se encontrarán críticas de la sociobiología especulativa y la psicología evolutiva en Philip Kitcher, *Vaulting Ambition: Sociobiology and the Quest for Human Nature*, Cambridge (Massachusetts), The MIT Press, 1987, y David J. Buller, *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*, Cambridge (Massachusetts), The MIT Press, 2005. En John E. R. Staddon, *Adaptive Dynamics: The Theoretical Analysis of Behavior*, Cambridge (Massachusetts), The MIT Press, 2001, se hallará el argumento de que el enfoque neurocientífico del comportamiento es prematuro, así como una defensa explícita del enfoque de la «caja negra». Judith Schlanger, *Les Métaphores de l'organisme*, París, Vrin, 1971, es una investigación sobre las metáforas biológicas en el estudio de la sociedad.

Capítulo 15

Fisiología y neurociencia

La búsqueda de la base fisiológica (con frecuencia neurofisiológica) del comportamiento humano complejo ha hecho importantes avances en años recientes, sobre todo como consecuencia de nuevas técnicas de medición y observación. Hay pocas dudas de que esta línea de investigación tiene gran futuro, aun cuando algunos ejercicios actuales tal vez sean prematuros, toscos o especulativos. Mencionaré tres grupos de descubrimientos que parecen ser de especial pertinencia para los objetivos de este libro.

Miedo

En el Capítulo 8 afirmé que las *creencias* o los antecedentes cognitivos son el instrumento característico de desencadenamiento de las emociones. Me enfado si creo que tu tropiezo conmigo fue intencional, imprudente o negligente, pero no me molesta si creo que fue un accidente causado por el tropiezo de un tercero contigo o por un movimiento repentino del tren. En *La Prisionera* de Proust, los sentimientos de celos del narrador crecen y menguan en función de sus creencias sobre lo que Albertine puede haber urdido durante los períodos en que él la ha dejado alejarse. Podríamos preguntarnos, no obstante, si las emociones no son a veces causadas por meras *percepciones* que, a diferencia de las creencias, no se «refieren» a nada. Descartes pensaba, por ejemplo, que la sorpresa o el asombro pueden «ocurrir antes que sepamos de ninguna manera si este objeto nos es conveniente o no».

Los trabajos neurofisiológicos sobre el miedo (en ratas) confirman esta idea. Hay dos caminos diferentes desde el aparato sensorial en el tálamo hasta la amígdala (la parte del cerebro que genera tanto las respuestas viscerales como las respuestas emocionales de conducta). Como confirmación de la idea tradicional de que las emociones siempre son precedidas y desencadenadas por una cognición, un camino va del tálamo al neocórtex, la parte pensante del cerebro, y del neocórtex a la amígdala. El organismo recibe una señal, se forma una creencia acerca de su significado y luego reacciona emocionalmente. También hay, sin embargo, un camino directo desde el tálamo hasta la amígdala que soslaya por completo la parte pensante del cerebro. En comparación con el primer camino, el segundo es «rápido y sucio». Por un lado, es más veloz. En una rata, un estímulo acústico tarda doce milisegundos (doce milésimas de segundo) en llegar a la amígdala a través del camino talámico, y casi el doble a través del camino cortical. Por otro lado, el segundo camino diferencia con menor fineza las señales entrantes. Mientras que la corteza puede distinguir que una tenue forma curva en un sendero es un palo torcido y no una serpiente, la amígdala es incapaz de hacer esa distinción. No obstante, desde el punto de vista de la supervivencia, el coste de reaccionar ante un palo como si fuera una serpiente debe haber sido mucho más pequeño que el coste del error contrario.

No se sabe si estos descubrimientos hechos en el estudio del miedo pueden generalizarse a otras emociones. De manera conjetural, algo parecido podría suceder con la ira. Cuando quedamos expuestos a una situación que podría ser un ataque, el coste de oportunidad de esperar hasta averiguar si en efecto lo es quizá sea demasiado elevado. Es muy posible que la selección natural haya predeterminado una tendencia a «disparar primero y preguntar después». Si golpeo y luego compruebo que en realidad no era la víctima de un ataque, puedo no obstante inventar una historia para justificar mi comportamiento. Este mecanismo bastante sutil, vinculado con nuestra necesidad de autoestima y resumido en el proverbio «quien ha ofendido no puede perdonar», interactúa con un mecanismo neurofisiológico que compartimos con los animales que carecen de esa necesidad de autoestima. Es probable que éste sea el patrón de muchos descubrimientos de la fisiología y la neurociencia. Las reacciones casi automáticas que compartimos con otras especies tal vez estén sujetas a las interpretaciones y elaboraciones autorreferenciales que son exclusivas de

los seres humanos. Esas racionalizaciones no son triviales, pues pueden llevarnos a persistir en la agresión en vez de admitir que estábamos en falta.

Otro descubrimiento del mismo programa de investigación sugiere una alternativa a la teoría freudiana de la represión del recuerdo. A menudo, las víctimas de violaciones o los soldados que han combatido no tienen un recuerdo consciente de su experiencia. ¿Es la no rememoración (la necesidad de este término neutro resultará evidente dentro de poco) un proceso motivado, como creía Freud, o las motivaciones no tienen absolutamente ningún papel en él? La respuesta quizá gire en torno de las relaciones entre el estrés y la formación (o no formación) del recuerdo. Un estrés moderado fortalece el recuerdo del suceso estresante, pero un estrés fuerte y prolongado puede elevar el nivel de los esteroides adrenales hasta el extremo de afectar desfavorablemente el hipocampo (la parte del cerebro donde se forman los recuerdos conscientes). En vez de ser reprimidos, es posible que los recuerdos del suceso traumático ni siquiera se hayan formado. De ser acertada, esta conclusión no implica que el suceso traumático no deje ninguna huella psíquica. Supongamos que tengo un accidente automovilístico y el claxon se traba. Más adelante, el sonido de un claxon se convierte en un estímulo condicionado del miedo, va directamente a la amígdala y provoca las respuestas corporales que por lo común se dan en una situación de peligro. Una vez más, es una percepción y no una creencia la que desencadena la emoción del pánico.

Confianza

Anticipándome al análisis de la *confianza* en el Capítulo 21, querría mencionar tres descubrimientos experimentales mentalmente estimulantes. Todos ellos se basan en un juego de la confianza (JC) entre dos jugadores, un «inversor» y un «operador». El inversor recibe cierta dotación, que puede transferir en parte o en su totalidad al operador. El experimentador hace luego que la transferencia rinda frutos, de manera tal que el operador recibe multiplicada varias veces la suma transferida. Para terminar, el mismo operador decide si hace un «reembolso» al inversor y, de hacerlo, determina también hasta dónde llegará su generosidad. En algunos experimentos, el inversor tiene asimismo la opción de castigar al operador si considera poco generoso el reembolso. Si ambos jugadores son racionales y egoístas y cada uno sabe que el otro lo es, no habrá transferencias en

interacciones anónimas y únicas, pese a lo cual los niveles positivos de transferencias y reembolsos son la regla en los experimentos. Otra variante del JC se considerará con mayor detalle en el Capítulo 21. Aquí me limitaré a mencionar algunos descubrimientos que relacionan el comportamiento de los inversores con su estado hormonal y la activación de centros del placer en su cerebro.

El primer experimento estudia la magnitud de la inversión como una función de la presencia o ausencia de la hormona oxitocina. Ya se sabía que esta hormona estimula un comportamiento prosocial en roedores y promueve la liberación de leche en las hembras humanas, pero el descubrimiento de que también propicia el comportamiento prosocial o la confianza en los seres humanos fue una sorpresa. Tras recibir la hormona, el porcentaje de inversores que transfirieron el total de su dotación al operador pasó del 21% al 45%. Otros tres descubrimientos son curiosos. Primero, los operadores a quienes se administró la hormona no hicieron reembolsos más grandes. Segundo, los inversores que la recibieron tenían las mismas creencias sobre la confiabilidad de los operadores (es decir, expectativas sobre los reembolsos) que aquellos a quienes no se aplicó. Tercero, cuando los inversores sabían que los reembolsos eran generados por un mecanismo aleatorio con la misma distribución de retribuciones que cuando jugaban con una persona real, la oxitocina no influía en la magnitud de las transferencias.[1] La interpretación natural es que la influencia de la hormona sobre el comportamiento consistía en hacer que los inversores mostraran menos «aversión a la traición» y no al riesgo. Otros experimentos que no se apoyan en manipulaciones fisiológicas también confirman la importancia de la aversión a la traición.

El segundo experimento, que permitía a los inversores castigar a los operadores poco generosos, estudió lo que sucedía en su cerebro mientras aplicaban ese castigo. En este JC, el inversor tenía que elegir entre transferir toda su dotación de diez unidades monetarias (UM) al operador y no transferir nada. Si hacía la transferencia, el experimentador la cuadruplicaba, con lo cual el operador quedaba en posesión de una dotación total de cincuenta UM: la dotación original de diez más las cuarenta generadas por la inversión. El operador debía optar entonces entre reembolsar veinticinco de las cincuenta al inversor y no devolver nada. En otras palabras, los tres resultados posibles eran (10, 10), (25, 25) y (50, 0) (véase la Figura 15.1).

Además, luego de que el operador tomara su decisión, ambos juga dores recibían una dotación adicional de veinte UM. El inversor podía utilizar su nuevo ingreso para castigar al operador, en una de dos condiciones. En una condición «costosa», podía asignar hasta veinte «puntos de castigo» al operador; cada punto hacía que el primero perdiera una UM y el segundo, dos. Así, al aplicar el máximo castigo, el inversor podía asegurarse de que la retribución del operador se redujera de setenta (50 + 20) a treinta, mientras que la suya propia se reducía de veinte a cero.[2] En una condición «sin costes», el operador era el único afectado por el castigo.

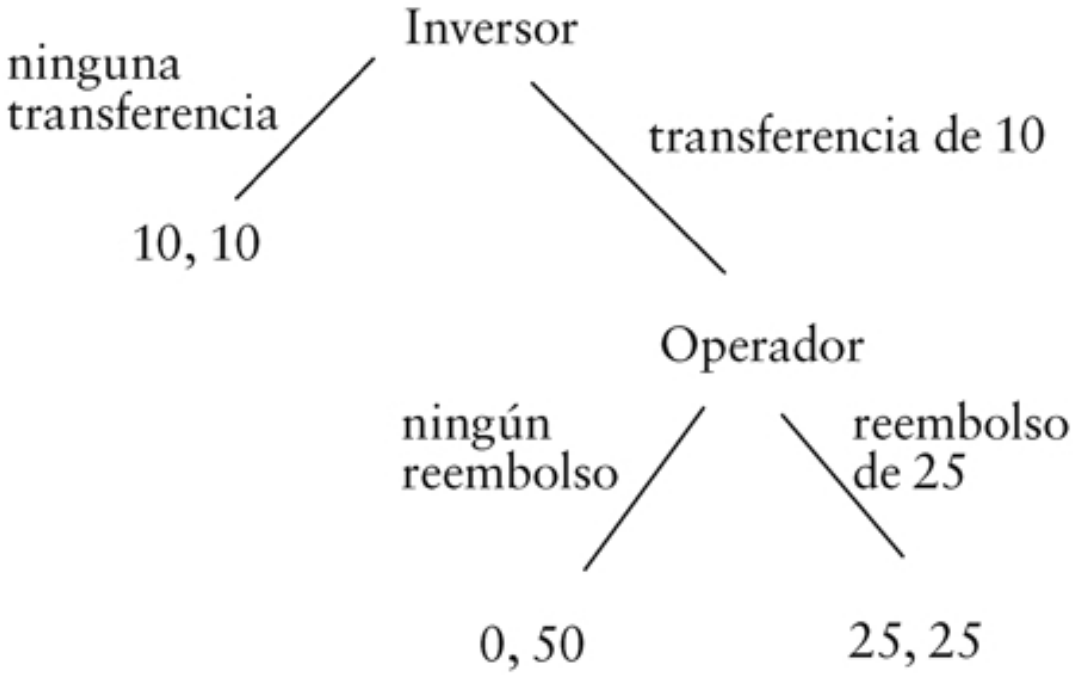


FIGURA 15.1

Con una sola excepción, los quince inversores hicieron transferencias de manera uniforme. El experimento se organizó de tal manera que cada uno de ellos jugara contra siete operadores, tres de los cuales hacían el reembolso, mientras los cuatro restantes conservaban todo lo recibido. Estos operadores egoístas eran el foco del experimento. Cuando un operador anunciaba su decisión de guardarse todo, el inversor tenía un minuto para reflexionar y decidir si quería castigarlo, y con cuánta severidad. Durante este lapso, se le exploraba el cerebro para detectar actividades en las

diversas regiones que podían ser relevantes. Una región, el núcleo caudado, está íntimamente vinculada al procesamiento de recompensas. Otra, la corteza prefrontal y orbitofrontal, está asociada a la integración de procesos cognitivos separados, por ejemplo las compensaciones entre costes y beneficios. En cada una de esas regiones, el patrón de actividades confirmó la hipótesis sobre la motivación para el castigo que voy a exponer a continuación.

Tanto en la condición costosa como en la condición sin costes, había una correlación entre la activación de los circuitos vinculados con las recompensas y el castigo monetario concreto que se imponía. Esta correlación podría significar o bien que la decisión de castigar genera satisfacción o bien que la satisfacción esperada del castigo induce la decisión de castigar. Para distinguir entre las dos hipótesis, el experimento consideró a once sujetos que habían impuesto el máximo castigo posible en la condición «sin costes». Entre ellos, aquellos cuyos circuitos de recompensas experimentaban la mayor activación también imponían los castigos más severos en la condición «costosa». Como la aplicación del castigo los complacía, estaban dispuestos a gastar más en ello, y de ese modo daban respaldo a la segunda hipótesis. También confirma esta interpretación el hecho de que la corteza fuera más activada en la condición costosa, cuando los sujetos tenían que compensar entre sí los costes materiales y los beneficios psíquicos del castigo, que en la condición sin costes.

Este hallazgo parece confirmar la teoría de la «satisfacción moral» de esta forma específica de comportamiento altruista (Capítulo 5). El castigo, si bien altruista en el sentido indirecto del comportamiento de beneficiar a terceros con algún coste material para el agente, no tiene su origen en motivaciones altruistas. Es demasiado pronto para decir si los resultados se mantendrán y si pueden generalizarse a formas directas de altruismo (dar dinero a los pobres) o al castigo de terceros (A castiga a B por perjudicar injustamente a C). La imparcialidad del castigo a terceros podría no generar la dulce satisfacción que proporcionan las represalias por una afrenta personal. Lo que parece claro, no obstante, es que la ciencia ofrece hoy los medios para decidir entre hipótesis motivacionales antagónicas que eluden la mera introspección. Si el castigo es motivado por el placer esperado de castigar, esta meta no puede ser el objetivo consciente del agente. Con excepción de los sádicos, la gente castiga porque cree que el otro *lo merece*.

Para tomar un ejemplo aún más inequívoco, es una incoherencia conceptual suponer que la motivación consciente de las personas para dar a los pobres puede ser *exclusivamente* la satisfacción moral generada por esa actitud, dado que dicha satisfacción es parasitaria de la creencia de que hacemos algo bueno para otros, no para nosotros mismos. Los estados que son inaccesibles a la conciencia podrían ser hoy accesibles a los escáneres cerebrales.

La *aplicación* de un castigo difiere de la *observación* de un castigo merecido. En un tercer experimento, se exploró el cerebro de un sujeto mientras éste observaba cómo se aplicaba una descarga dolorosa a dos personas (cómplices del experimentador) que habían interactuado con él en un juego de la confianza, una de las cuales había mostrado un comportamiento limpio, en tanto que la otra se había comportado injustamente. Cuando el sujeto observaba el castigo administrado al jugador limpio, la exploración cerebral mostró la activación de las áreas del cerebro relacionadas con el dolor («siento tu dolor»). La observación del castigo a los jugadores injustos causaba una menor activación de dichas áreas, y, en cambio, una mayor activación de las áreas relacionadas con el placer (*Schadenfreude*). (Un descubrimiento importante, pero no explicado, fue que las mujeres sentían más dolor y menos placer que los hombres cuando se castigaba a los sujetos injustos.) El *Schadenfreude* es una especie de ira sucedánea: A se complace al observar a B castigar a C cuando C ha perjudicado a A. Sería interesante investigar las actividades cerebrales en la indignación sucedánea: ¿se complacería tanto A al observar a B castigar a C cuando C ha perjudicado a D? También podría compararse la ira sucedánea con la ira directa: ¿cómo se compara el placer de observar a B castigar a C cuando C ha perjudicado a A con el placer de A al castigar a C cuando C ha dañado A? Estas emociones elementales, bien conocidas gracias a la ficción, el teatro y la poesía, parecen estar hoy al alcance de la ciencia.

El completado

Los lectores que utilizan correctores ortográficos sabrán que el programa es capaz de «adivinar» o «completar» lo que podría faltar en una palabra mal escrita. A veces, el resultado es una palabra equivocada, que (de no ser corregida) puede tener efectos cómicos y hasta desastrosos, pero en términos generales el programa es una herramienta confiable y útil.

También el cerebro tiene la capacidad de llenar las lagunas de la experiencia. Una figura circular con pequeñas aberturas se verá como un círculo completo o cerrado. Si las aberturas son muchas, tal vez no lo veamos literalmente como un círculo completo, pero de todos modos advertiremos que las partes visibles se encuentran en un círculo invisible. En la parte superior de la Figura 15.2, «vemos» los círculos debajo de los rectángulos, pero cuando éstos se sacan, como en la parte inferior, los círculos también desaparecen.

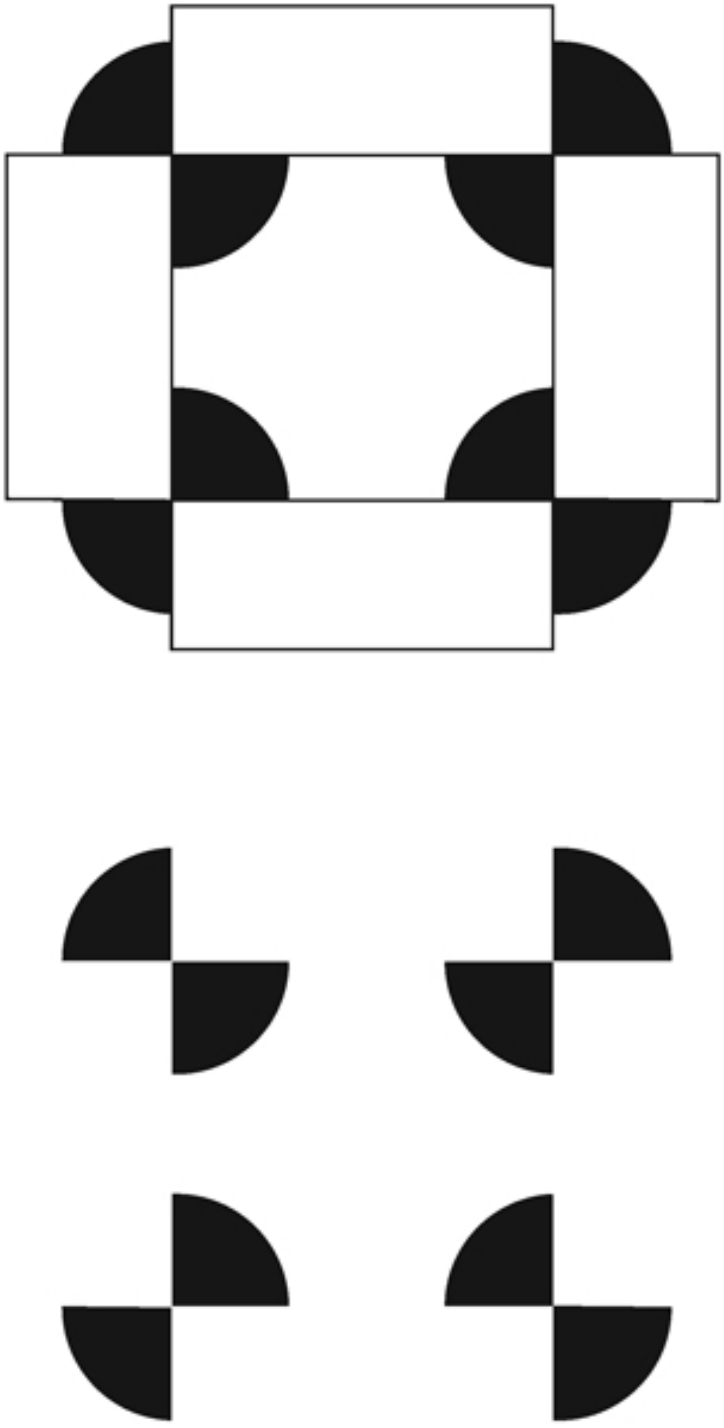


FIGURA 15.2

El completado es un fenómeno cotidiano. El otro día caminaba por Levallois-Peret, en las afueras del territorio propiamente dicho de París, y vi el letrero de una calle que leí como «*rue* Auguste Blanqui». Me pregunté por qué, en ese rico barrio presuntamente conservador, se daría a una calle el nombre del más violento revolucionario francés del siglo XIX. Cuando miré con más detenimiento, me di cuenta que decía «*rue* Aristide Briand», en referencia a un estadista francés de credenciales impecables. Lo que había pasado, claro está, es que mi mente vio la primera letra del nombre y el apellido, así como la longitud aproximada de uno y otro, y, como conozco mucho más a Blanqui que a Briand, inferí que veía un letrero que lo conmemoraba. No interpreté el letrero completando conscientemente las letras faltantes de una manera u otra: el completado se produjo en forma inconsciente, y mi experiencia consciente consistió en *ver* las dos palabras completas. En otra oportunidad, buscaba una panadería (*boulangerie*) y reparé en el letrero de una tienda con esa designación, para descubrir a continuación que lo que había visto era la segunda mitad de «Restaurant de l'Orangerie». A diferencia del caso de Blanqui y Briand, éste bien puede haber sido un caso de *percepción errónea motivada*, y no un ejemplo de «hipótesis más verosímil» del cerebro sobre lo que veía.

La actitud de plantear la hipótesis más verosímil también podría denominarse «tendencia a sacar conclusiones precipitadas». Para nuestra fortuna, el cerebro lo hace todo el tiempo. Los mecanismos de completado, como los correctores ortográficos, son útiles; indispensables, a decir verdad. Sin embargo, a veces la búsqueda de patrones se convierte en una búsqueda obsesiva de significados. El cerebro es un teórico natural de conspiraciones; para decirlo con mayor exactitud, según algunas teorías, el hemisferio cerebral derecho tiene la función de imponer un marco coherente a todas las informaciones que nos bombardean constantemente. Esto podría explicar algunos casos de «anosognosia» o negación de la enfermedad por quienes la sufren. En vez de ser un fenómeno *motivado*, como lo sugeriría la explicación freudiana, la negación podría ser simplemente la hipótesis más verosímil del corrector ortográfico del cerebro. Más probablemente, al parecer, el fenómeno podría deberse a mecanismos tanto motivacionales como no motivacionales.

Sospecho que las actividades cerebrales de «completado» y «atribución de significados» están presentes no sólo en los sujetos estudiados por los científicos, sino también en muchos investigadores. No digo que no haya

diferencias entre las teorías conspirativas de Hamas, por un lado, y la crítica literaria posmoderna, la explicación funcional o el psicoanálisis, por otro; sólo digo que todos parecen animados por la necesidad de encontrar significado y coherencia, más allá de lo que pueden justificar los hechos en cuestión. En el caso de las hipótesis científicas, los atajos y las extrapolaciones no pueden justificarse ni explicarse por la necesidad de tomar medidas inmediatas, como es habitual hacerlo en la vida diaria. Reflejan, por el contrario, el profundo desasosiego que sentimos cuando somos incapaces de atribuir un sentido al mundo, ya se nos requiera o no actuar en función de éste. La necesidad de un cierre cognitivo y la intolerancia a la admisión de ignorancia (Capítulo 7) también reflejan ese desasosiego, al igual que lo hace la hiperracionalidad exhibida por quienes, para tomar una decisión racional, invierten más de lo que está en juego en la decisión misma (Capítulo 12). La tendencia a «culpar a la víctima» es otro ejemplo. Al suponer que el mundo es fundamentalmente justo, de modo que la gente «recibe lo que se merece», tendemos a desvalorizar y menospreciar a las víctimas de sucesos puramente azarosos, aun la selección por sorteo para el servicio militar; a decir verdad, incluso las víctimas mismas tienden a hacerlo.

* * *

NOTA BIBLIOGRÁFICA

El examen del miedo se apoya en Joseph E. LeDoux, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, Nueva York, Simon & Schuster, 1996 [*El cerebro emocional*, Barcelona, Ariel, 1999]. El impacto de la oxitocina sobre la confianza se muestra en Michael Kostfeld y otros, «Oxytocin increases trust in humans», *Nature*, 435, 2 de junio de 2005, págs. 673-676. La idea de aversión a la traición es confirmada por Iris Bohnet y Richard Zeckhauser, «Trust, risk and betrayal», *Journal of Economic Behavior and Organization*, 55(4), 2004, págs. 467-484. El estudio de la confianza y la venganza es de Dominique J. F. de Quervain y otros, «The neural basis of altruistic punishment», *Science*, 305(5688), agosto de 2004, págs. 1254-1258. El estudio de la confianza y la *Schadenfreude* es de Tania Singer y otros, «Empathic neural responses are modulated by the perceived fairness of others», *Nature*, 439, 26 de enero de 2006, págs. 466-469. Una guía estimulante sobre los fenómenos de

«completado» es Vangipuram S. Ramachandran y Sandra Blakeslee, *Phantoms in the Brain: Human Nature and the Architecture of the Mind*, Nueva York, Quill, 1998 [*Fantasmas en el cerebro: los misterios de la mente al descubierto*, Madrid, Debate, 1999], que señala que la hostilidad inicial de Ramachandran a las explicaciones freudianas de la anosognosia fue reemplazada por la creencia de que debe estar presente un elemento de la negación inconsciente; véase también Neil Levy, «Self-deception without thought experiments», en au.geocities.com/neil_levy/Documents/articles/Self-deception.pdf. Los descubrimientos sobre la «culpa de las víctimas» figuran en Zick Rubin y Anne Pepau, «Belief in a just world and reaction to another's lot: a study of participation in the national draft lottery», *Journal of Social Issues*, 29, 1973, págs. 73-93.

Notas al pie

[1] Las retribuciones reales de los inversores en juegos de la confianza contra un operador real se utilizaron para determinar las distribuciones de las recompensas que se les indica esperar cuando juegan contra un mecanismo aleatorio. Suponiendo que los sujetos del JC tienen *expectativas racionales*, es decir creencias acertadas sobre las distribuciones de las recompensas que pueden contar con recibir, en ambas condiciones están frente a las mismas distribuciones. Este supuesto podría cuestionarse.

[2] En las sumas realmente pagadas a los sujetos, cada punto correspondía a 0,1 franco suizo. La suma parece insignificante, pero otros experimentos muestran que en juegos de este tipo se dan los mismos resultados cuando las posturas alcanzan el nivel de un salario mensual. En su mayor parte, estos resultados se obtienen utilizando subsidios de investigación otorgados por el Primer Mundo para llevar a cabo experimentos en países del Tercer Mundo.

Capítulo 16

Explicación por las consecuencias y selección natural

Refuerzo

Toda explicación es causal. Para explicar un suceso, mencionamos su causa. Las causas preceden a sus efectos en el tiempo. De ello se sigue que no podemos explicar un suceso, como una acción, por sus consecuencias. Sin embargo, si el *explanandum* es un *patrón* de comportamiento recurrente, las consecuencias de ese comportamiento en una oportunidad pueden ser parte de las causas que hacen más probable su aparición en una oportunidad ulterior. Esto puede suceder principalmente de dos maneras: por *refuerzo* y por *selección*. Me concentraré en la segunda, ya que es más importante para mi objetivo, pero comenzaré con algunas palabras sobre el primero.

Si las consecuencias de un comportamiento dado son agradables o gratificantes, tendemos a repetirlo más a menudo; si son desagradables o punitivas, aparecerá con menos frecuencia. El mecanismo subyacente podría ser simplemente la elección racional consciente, si *advertimos* las consecuencias agradables o desagradables y *decidimos* actuar en el futuro con el fin de repetir o evitar repetir la experiencia.[3] Muchas veces, con todo, el refuerzo puede producirse sin una elección intencional. Cuando los lactantes aprenden a llorar porque los padres los levantan si lo hacen, no hay razón para pensar que en primer lugar notan los beneficios de llorar y

luego lloran a voluntad para obtenerlos. Cuando niños más crecidos hacen un berrinche para salirse con la suya, los padres suelen decir que su enojo no es auténtico.

El aprendizaje por refuerzo ha sido ampliamente estudiado en experimentos de laboratorio con animales. Uno de los más característicos brinda al animal la oportunidad de presionar una palanca, o una entre varias, y recompensa las presiones sea como función de la cantidad habida desde la última recompensa, sea como función del tiempo transcurrido desde ésta. En uno y otro caso, la función puede ser determinista o probabilística. En los esquemas de *proporción fija*, el animal recibe una recompensa luego de haber presionado una palanca una cantidad fija de veces, mientras que en los esquemas de *proporción variable* el número de presiones necesarias para recibirla varía en forma aleatoria. En ambos casos, cada presión de la palanca genera un «punto de recompensa» que se suma a los anteriores. En los esquemas de *intervalo fijo* una presión producirá una recompensa determinado tiempo después de haberse entregado la anterior, mientras que en los esquemas de *intervalo variable* el tiempo que debe pasar antes del ofrecimiento de una nueva recompensa varía al azar. En cualquiera de los dos casos, el momento de la recompensa es independiente de la cantidad de presiones. Cada esquema de refuerzo produce, luego de algún tiempo de aprendizaje, un patrón específico y estable de comportamiento que, por otra parte, desaparecerá del patrón correspondiente una vez eliminado el reforzador (la recompensa). Por ejemplo, las respuestas que se aprenden por medio de una recompensa otorgada a cada presión de la palanca (un caso especial del esquema de proporción fija, conocido como refuerzo continuo) desaparecen con mayor rapidez que las aprendidas en un esquema de proporción variable aleatoria. La intuición quizá sugiriera lo contrario, dado que el refuerzo continuo parecería producir un hábito más fuerte, pero, como a veces pasa, la intuición se equivoca.

La relevancia de estos descubrimientos fuera del laboratorio depende de los objetivos. Si la meta es *dar forma* a la acción, pongamos por caso, en una situación áulica, un casino o el lugar de trabajo, un diseñador puede (de manera más o menos libre) imponer un esquema de recompensas para generar el comportamiento deseado. Por ejemplo, los esquemas de intervalo variable se utilizan a menudo para modelar la conducta, como ocurre cuando un docente anuncia una política de exámenes al azar. En el esquema

de proporción variable que funciona en muchos juegos de apuestas, es más fácil establecer el comportamiento si la primera recompensa aparece en los momentos iniciales.[4] Como los gerentes de casinos y pistas de carreras carecen de la tecnología para absorber novatos mediante el ofrecimiento de grandes ganancias, tienen que contar con el hecho de que por las leyes del azar, algunos jugadores tendrán la suerte del principiante.[5] Sin embargo, las operaciones de los timadores se apoyan a menudo en el aliciente deliberado de ganancias tempranas para el incauto. En el aula y el casino, los esquemas de recompensas actúan «a espaldas» de los estudiantes o los jugadores, en el sentido de que no dan forma al comportamiento por medio de incentivos explícitos sino, como en el caso del lactante llorón, a través de un proceso inconsciente. En contraste, cuando los administradores pagan a los empleados una vez que éstos alcanzan un objetivo determinado (esquema de proporción fija) o por mes (esquema de intervalo fijo), no hacen sino establecer un sistema de incentivos. Como el comportamiento de los empleados puede explicarse adecuadamente con referencia a la recompensa *esperada*, no hace falta apelar a la recompensa *real*.

Cuando la meta consiste en *explicar* los patrones de conducta por sus consecuencias concretas, los esquemas de recompensas sólo son relevantes si aparecen naturalmente y, por añadidura, son lo bastante opacos para no crear incentivos explícitos. Esto no parece suceder con frecuencia en el caso de los dos esquemas fijos. En la vida cotidiana, el mero número de respuestas muy pocas veces es decisivo para la recompensa. Lo que da forma al comportamiento que mis amigos tienen conmigo no es la cantidad de sonrisas amistosas que les prodigo, sino la coherencia y la adecuación de mi sonreír. En los ámbitos naturales, las recompensas que llegan cada tanto tiempo, como el cheque de mi salario, son contadas. Los dos esquemas variables son más importantes. Una persona que juega a «acercarse y alejarse» (un esquema de proporción variable) con un miembro del sexo opuesto (o del mismo) tal vez induzca una atracción más fuerte que quien exhibe de manera invariable un comportamiento amistoso. Un esquema de intervalo variable surge cuando tratamos de comunicarnos con alguien por teléfono y la línea está ocupada. Sabemos que, si volvemos a llamar, tarde o temprano nos comunicaremos, pero no sabemos cuándo. Esta situación genera un patrón de llamadas uniformes que no sería la predicción única de la teoría de la elección racional. Esa teoría podría predecir un número indefinido de patrones, según cuál fuera la creencia de quien llama acerca

de la duración probable de la conversación telefónica de la otra persona. Parece improbable, sin embargo, que la gente tenga creencias estables sobre asuntos como ese.

El patrón de respuestas generado por el refuerzo no es, en general, el que produciría la elección racional y consciente. Supongamos, por ejemplo, que un animal tiene la opción de presionar una u otra de dos palancas, una de las cuales recompensa según un esquema de proporción variable, y otra que lo hace de acuerdo con un esquema de intervalo variable. El patrón racional, que maximizará la recompensa global, consiste en presionar la mayoría de las veces la palanca de la proporción variable, para acumular puntos de recompensa, y acudir a la palanca del intervalo variable de vez en cuando para ver si ha aparecido una nueva recompensa. No es ése, sin embargo, el patrón producido por el aprendizaje por refuerzo. En cambio, los animales presionan la palanca del intervalo variable mucho más a menudo de lo que sería óptimo. Al hacerlo, igualan las recompensas *promedios* a la presión de una u otra palanca, en vez de igualar las recompensas *marginales*, como dictaría la racionalidad. En el caso de otras combinaciones de esquemas, el aprendizaje por refuerzo remeda en ocasiones la elección racional, pero no de una manera consistente. Si hay algún mecanismo no intencional capaz de simular en forma confiable la racionalidad, tendremos que buscarlo en otra parte.

Aptitud reproductiva diferencial

El mecanismo más frecuentemente mencionado es el de la *selección* natural o social. En la tercera parte de este libro he analizado cómo podemos explicar el comportamiento si suponemos que *los agentes se adaptan a su medio ambiente*, de manera más o menos racional. En una perspectiva radicalmente diferente, podemos tratar de explicarlo suponiendo que *el medio ambiente selecciona a los agentes*. Aunque la selección puede ser obra de un agente intencional, como ocurre cuando se cría a los perros domésticos con el fin de que sean dóciles o a las ratas de laboratorio para que sean más inteligentes, muchos mecanismos de selección se basan en procesos causales que no implican la participación de ningún agente de esas características.

En particular, la *supervivencia diferencial* de organismos sobre la base de sus patrones de conducta puede llevar a un comportamiento óptimo

(óptimo para la reproducción) en la población aun en ausencia de toda elección o intención optimizadoras. Supongamos que el 10% de los organismos en una población de cien integrantes se alimentan de manera tan eficiente que tienen en promedio diez descendientes que sobreviven hasta la adultez, mientras que el promedio del 90% restante es de sólo cinco. Si el comportamiento de los padres se transmite (por cualquier mecanismo) a la descendencia, la siguiente generación de organismos adultos incluirá una fracción de 100/550 = 2/11 ~ 18% que exhibirá el comportamiento más eficiente. Con el paso de algunas generaciones más, virtualmente todos los organismos desplegarán ese comportamiento. Si preguntamos *por qué* está universalmente presente, la respuesta es que tiene mejores consecuencias.[6] Este mecanismo funciona a través de las generaciones. A diferencia del aprendizaje por refuerzo, no modifica el comportamiento de ningún individuo determinado, sólo el comportamiento típico de generaciones sucesivas de individuos.

Selección natural

La teoría de la selección natural expone con gran detalle esta historia. Aquí presentaré una versión simplificada o «clásica» de la teoría, que basta para lo que me propongo. Quiero mostrar, en particular, que la selección natural difiere de la elección intencional y, a fortiori, de la elección racional. La selección natural es un mecanismo de optimización, pero sólo en un sentido débil. También quiero sentar las bases del argumento que expondré en el próximo capítulo, a saber, la improbabilidad de que la selección *social* produzca siquiera esta forma débil de optimización.

La aptitud de un organismo, medida por la cantidad de sus descendientes (en condiciones de reproducirse),[7] es determinada en forma conjunta por su medio ambiente y sus propiedades fisiológicas y de conducta, o *fenotipo*. Su *genotipo* es la serie de instrucciones que, también en forma conjunta con el medio ambiente, determinan el fenotipo. Esas instrucciones están codificadas en moléculas largas de ADN, que tienen muchas propiedades en común con un lenguaje escrito. Las letras del ADN son cuatro moléculas (nucleótidos) llamadas T, A, G y C. Las palabras o *codones* del ADN son tripletes de nucleótidos; cada uno de esos tripletes contiene instrucciones para armar uno de los veinte aminoácidos que son los elementos constituyentes fundamentales de las proteínas en el

organismo. Como hay sesenta y cuatro tripletes (sesenta y uno si excluimos tres que tienen una función diferente), más de uno de ellos puede codificar el mismo aminoácido. Un *gen* es un segmento de ADN que codifica una proteína específica.

Los organismos que se alimentan con mayor eficiencia aparecieron como resultado de una mutación azarosa del genoma. En la imagen clásica de la evolución, se suponía que las mutaciones eran la resultante de pequeños errores aleatorios en la duplicación del material genético ocurrida durante el proceso de reproducción. Hay varios tipos de pequeñas mutaciones: supresión de un nucleótido, inserción de un nucleótido o sustitución de un nucleótido por otro. En esta exposición simplificada, sólo me referiré al tercer tipo, el más común. Aunque hoy se interpreta que los mecanismos como la duplicación de genes pueden producir macromutaciones, por ahora me atendré a la imagen clásica.

En el marco de esa imagen, el aporte de las mutaciones a la evolución puede ilustrarse por medio de una analogía con la estructura de tres niveles de una oración impresa. La Figura 16.1 muestra las similitudes entre los dos casos.

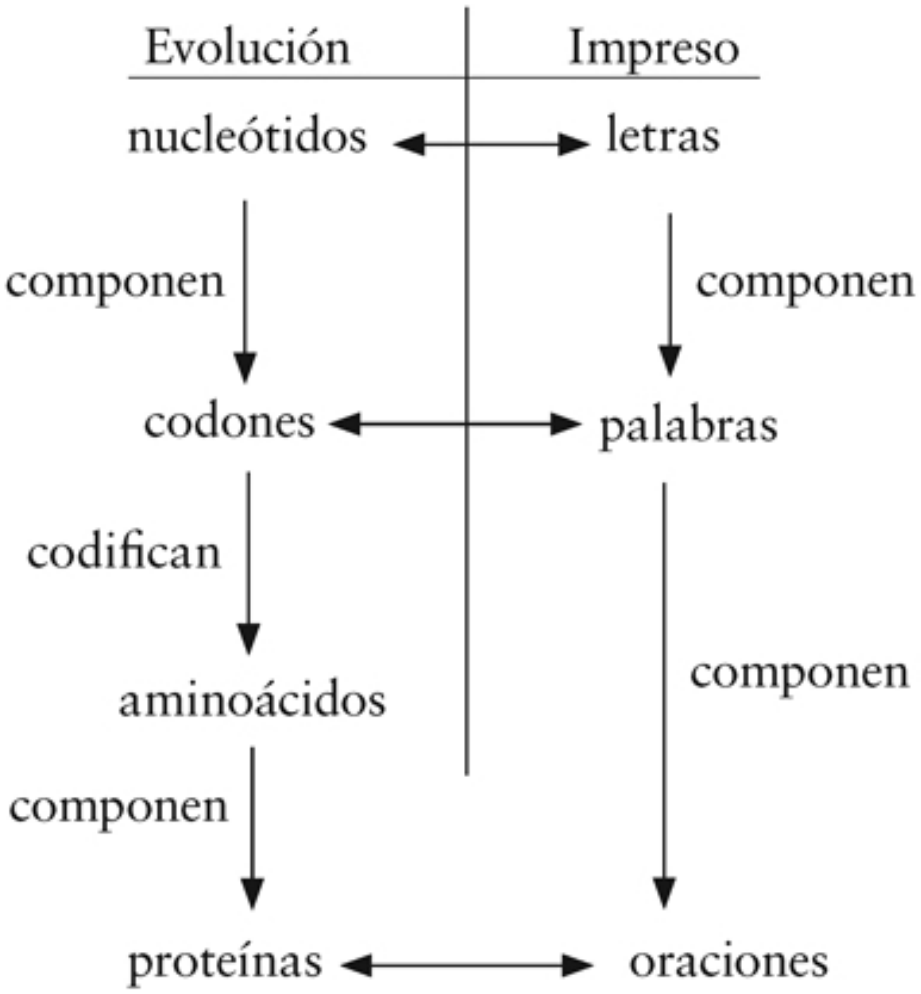


FIGURA 16.1

Cuando un libro vuelve a componerse sobre la base de la primera edición, una escasa concentración del tipógrafo puede dar lugar a algunas desviaciones con respecto al texto original. Podemos concebirlas como la sustitución de una letra por otra, que, por ejemplo, convierte «mano» en «mago» o «bano». Los errores son azarosos en el sentido específico de que carecen de relación con el contenido del libro. De manera similar, las mutaciones son aleatorias en el sentido de que su probabilidad de aparición no está relacionada con la naturaleza de los cambios fenotípicos que provocarán. Algunas letras pueden cambiarse más que otras, si el tipógrafo tiene dificultades para distinguir la «m» de la «n». La segunda edición de algunos libros quizá contenga más errores que otras, a raíz, por ejemplo, de

que el tipógrafo trabajaba borracho. Análogamente, los agentes mutagénicos pueden afectar las velocidades de mutación.[8]

De ordinario, los errores de imprenta en un libro o en cualquier mensaje escrito tienen un impacto negativo sobre el contenido, a veces (literalmente) fatal. El desplazamiento de una coma, por ejemplo, puede ser un asunto de vida o muerte.[9] De manera similar, la mayoría de las mutaciones tienen consecuencias nocivas para el organismo en el cual ocurren, en cuanto reducen su aptitud reproductiva. De vez en cuando, sin embargo, puede producirse una mutación que aumenta la aptitud del organismo, así como un fracaso accidental en el intento de reproducir el texto de la primera edición podría suministrar una información más precisa o actualizada. No obstante, en el siguiente paso del argumento la analogía se derrumba. Una mutación favorable va a estar representada con mayor abundancia en la generación siguiente, dado que el organismo en el cual ha ocurrido tendrá más descendencia, y las mutaciones se transmiten a ésta. En contraste, no hay mecanismo en virtud del cual un libro que corrige accidentalmente los errores de la primera edición tienda a venderse mejor.

En los errores tipográficos, cualquier letra puede ser reemplazada por cualquier otra. No es obligatorio que la nueva palabra tenga un significado; es tan probable que «mano» sea reemplazado por «bano» como por «mago». En las mutaciones, del mismo modo, cualquier otro nucleótido puede reemplazar cualquiera de los nucleótidos de un triplete. En contraste, no cualquier aminoácido puede surgir de cualquier otro por obra de la mutación de un solo nucleótido. Algunos cambios de proteínas pueden ocurrir de esta manera, pero otros no. Siempre es posible, sin embargo, ir de un aminoácido a otro por mutaciones puntuales sucesivas. El proceso se asemeja un poco al juego familiar en el cual una palabra determinada se transforma en otra mediante reemplazos de una letra, siempre que todas las palabras intermedias estén en el diccionario. Se puede ir de PELO a PESO en un paso y de PELO a MITO en cuatro (PELO-PETORETO-RITO-MITO), pero estoy bastante seguro de que no hay pasos suficientes para llevarnos de PELO a ÓRIX, a menos que nos desviemos del diccionario. En el «juego de las mutaciones» correspondiente, los sesenta y cuatro codones están «en el diccionario», excepto tres que actúan como «períodos» (señales de terminación de la síntesis proteica). Aunque estos tres «codones de *stop*» impiden algunos cambios de un paso en los tripletes, la restricción es demasiado débil para bloquear los cambios de varios pasos. Así, se puede ir

del aminoácido metionina al triptófano en dos pasos, sea de ATG a TTG (leucina) y de ahí a TGG o de ATG a AAG (glicina) y de ahí a TGG.[10] Además, hay caminos que implican más de dos pasos, como ATG-ACG-TCG-TGG.

Imaginemos ahora un juego familiar un poco más complicado que el que se juega habitualmente y supongamos que cada palabra está inserta en una frase; la palabra puesta en lugar de otra no sólo debe estar en el diccionario, sino producir una oración con significado.[11] Ahora bien, los reemplazos aceptables de letras tienen dos restricciones. Supongamos que la frase inicial es «bajo el nido». Si la reemplazo por «rajo el nido», respeto ambas restricciones, la del diccionario y la del significado. Al reemplazarla por «fajo el nido», sólo respeto la restricción del diccionario. Si la reemplazo por «fajo el nipo», no respeto ninguna de las dos. Y si cambio una vez más «fajo el nido» por «fajo el niño», cumplo con ambas restricciones. Para llegar a esa oración (con significado) desde la frase inicial (con significado) mediante sustituciones de una letra, tuvimos que apartarnos del conjunto de frases con significado.

La selección natural genera máximos locales

La analogía biológica es clara. Un aminoácido resultante de la sustitución de un nucleótido se inserta en una proteína. Como las proteínas son vitales para el organismo, tienen que ser biológicamente viables (el análogo de la significación para las oraciones). Si alguna proteína intermedia en el camino de la forma inicial a la forma final es inviable, no puede fijarse en la población, *dado que el organismo en el cual se produjo por primera vez la mutación de esa proteína no dejará ningún descendiente que pueda dar el siguiente paso*. Aun cuando la forma final de la proteína mostrara mayor aptitud en comparación con la de la primera etapa, ese hecho no generaría ninguna presión evolutiva en las etapas intermedias. Si *todos* los caminos factibles desde la primera hasta la última etapa exigieran al organismo, en algún punto, la adopción de la estrategia indirecta de «un paso atrás, dos pasos adelante», la forma final nunca podría alcanzarse. Como la mayoría, probablemente la abrumadora mayoría, de las mutaciones son deletéreas, esta situación es muy verosímil. En una forma simplificada (que omite muchos caminos), la estructura se muestra en la Figura 16.2. Tres proteínas diferentes confieren diferentes grados de aptitud (medida

verticalmente). La estructura del código genético permite mutaciones de un solo nucleótido de A a B y de B a C, pero no de A a C.

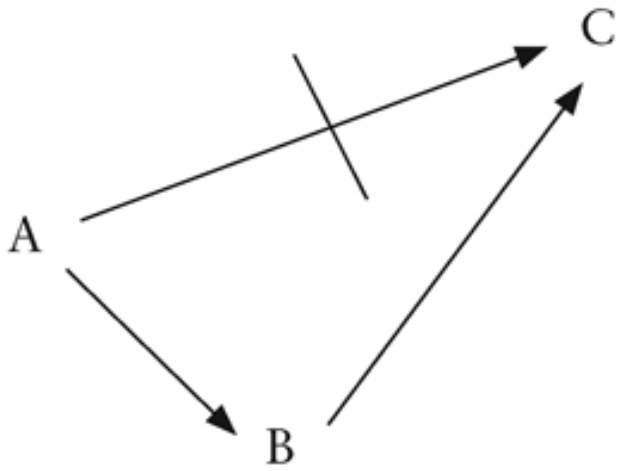


FIGURA 16.2

La selección natural (en esta imagen clásica) se ve limitada a pequeñas mejoras acumulativas. El organismo sube a lo largo de un gradiente de aptitud hasta que llega a un *máximo local*, definido como un estado en el cual todos los cambios adicionales de un paso reducirían la aptitud. Si bien puede haber picos más elevados en el «paisaje adaptativo», no será posible alcanzarlos mediante cambios de un solo paso. Este proceso difiere de la elección intencional en tres aspectos. Recuérdese (Capítulo 6) que, en virtud de su intencionalidad, los seres humanos son capaces de 1) valerse de estrategias indirectas, 2) esperar, y 3) apuntar por delante de un blanco móvil. Acabamos de ver que la selección natural es incapaz de 1). En cuanto a 2), considérese la Figura 16.3.

La estructura del código genético permite mutaciones de un solo nucleótido de A a B y de A a C, pero no de B a C. Si se produce una mutación de A a B, la población queda estancada en el máximo local B (de bajo nivel), puesto que están bloqueadas las mutaciones de un paso al máximo global C. No hay un mecanismo que favorezca una mutación a C en vez de a B. A diferencia de los agentes intencionales, la selección natural es *oportunista* y aprovecha cualquier ocasión de mejora, al mismo tiempo que es incapaz de *esperar* hasta la aparición de la ocasión óptima.

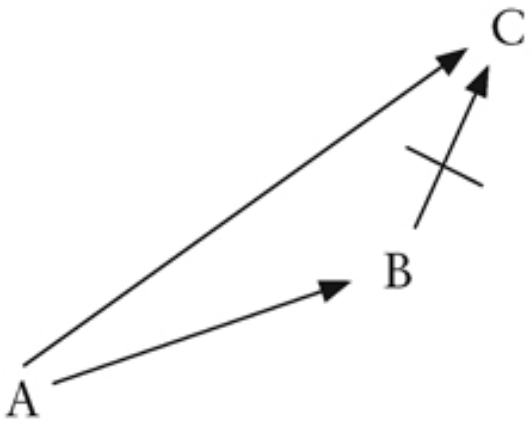


FIGURA 16.3

En lo que respecta a 3), las poblaciones se adaptan a un medio ambiente que está en constante cambio. Si los cambios son regulares, por ejemplo estacionales o diurnos, aquellas se adaptan a ellos. Si se produce un suceso único, como un repentino cambio climático, el comportamiento que con anterioridad se situaba en el máximo local de aptitud puede llegar a ser menos que óptimo, de modo que las mutaciones que previamente habrían sido deletéreas ahora resultan favorables. Si el cambio es prolongado, como ocurre cuando el clima se enfría o se calienta durante un extenso período, es posible que este proceso nunca llegue a un nuevo máximo local. La población explorará los cambios en el medio ambiente con una eficacia que dependerá de la velocidad relativa de los dos procesos. Las adaptaciones sorprendentemente finas que se observan en animales y plantas sugieren que éstos se ajustan al medio ambiente mucho más rápido de lo que este último cambia. No obstante, los organismos siempre quedarán un tanto rezagados, dado que no pueden *prever* los cambios en el medio ambiente. En contraste, los hombres pueden cobrar conciencia de cambios futuros como el calentamiento global y tomar precauciones contra ellos antes de que efectivamente ocurran o, si son la resultante del comportamiento humano, impedir que sobrevengan.

El medio ambiente de una población está constituido, entre otras cosas, por poblaciones de otras especies con las cuales aquella puede mantener una relación de presa y depredador. Como presa, tal vez desarrolle mejores estrategias de escape; como depredador, mejores estrategias de caza. Así como el individuo zorro y el individuo liebre se persiguen uno a otro a

través de los campos, las especies Zorro y Liebre se persiguen a lo largo de las generaciones. Pero mientras que la lógica de la selección natural impide al Zorro prever dónde va a ir la Libre dentro de algunos milenios, algunos depredadores son capaces de interceptar la trayectoria de huida de la presa. De manera similar, los procesos localmente maximizadores de la selección natural han generado la capacidad de maximización global que se constata en los seres humanos.

Las mutaciones son tanto un insumo como un producto de la selección natural, pues los efectos sobre la aptitud pueden establecer la velocidad a la que ocurren. Participan en ello varios mecanismos. En una población bacteriana constantemente expuesta a nuevos antibióticos, quizá se vean favorecidas las cepas con velocidades más elevadas de mutación. Éste es un caso especial de la idea más general de que en un medio ambiente cambiante, las poblaciones «inmutables» se extinguirán. Al mismo tiempo, las velocidades muy altas de mutación pueden conducir a la pérdida de valiosa información genética y a la extinción. En el medio, hay una velocidad óptima. En un medio ambiente constante, hay una compensación entre los beneficios de una baja velocidad de mutación y el coste de los mecanismos de «corrección de pruebas» y reparación que son necesarios para mantenerla.

Esta «imagen clásica» es simplista en varios aspectos que deberían señalarse si éste fuera un libro de texto. Como no lo es, me limitaré a afirmar que la imagen más compleja modifica pero no invalida la idea de que la selección natural tiende a estancarse en una «trampa de máximo local». Se producen grandes mutaciones, y algunas de ellas quizá sean responsables de transformaciones que no habrían sido posibles por medio de pequeñas mutaciones puntuales. Además, las formas inferiores no son eliminadas al instante por la competencia. En la Figura 16.2, la mutación a B no produce necesariamente un organismo inviable en el sentido estricto de ser incapaz de sobrevivir o reproducirse. Algunos organismos en el estado B podrían sobrevivir y producir organismos del estado C. En la Figura 16.3, algunos organismos en el estado A podrían sobrevivir a la competencia de los organismos más eficientes del estado B el tiempo suficiente para que se produjera una mutación al estado C. La probabilidad de aparición del máximo local tiene que ver con la velocidad relativa de dos procesos: la extinción de variedades inferiores y el índice de surgimiento de mutaciones favorables. No existe, sin embargo, ningún mecanismo que

pueda remedar, *de manera sistemática*, la capacidad de los seres intencionales de anticiparse, esperar o utilizar estrategias indirectas.

La tendencia de la selección a producir un comportamiento optimizador o estructuras óptimas a través del ascenso de un gradiente no debe llevarnos a concluir que todos los rasgos observados de los organismos son óptimos. Por un lado, como ya se ha mencionado, podría haber un *retraso*, si la población no se ha puesto a la par de los cambios en el medio ambiente. Lo que observamos tal vez haya sido óptimo en algún momento del pasado, pero ya no lo es. Por otro lado, el fenómeno de la *pleiotropía* nos dice que la selección natural puede favorecer características que son menos que óptimas si se las considera aisladamente, pero que en conjunto forman parte de una solución óptima. Si esto sucede, es porque un gen puede tener más de un efecto sobre el fenotipo. Si los efectos positivos predominan sobre los efectos negativos, la selección natural favorecerá el gen en cuestión. Como ejemplo, considérese la existencia de rasgos machos en insectos que causan daños físicos a sus parejas durante la copulación. Se ha sugerido que esos daños surgen o bien como un efecto secundario pleiotrópico negativo de la adaptación que da a los machos una ventaja reproductiva en otro contexto, o bien como una adaptación masculina en sí misma. Los experimentos indican que la primera explicación es más verosímil.

Las unidades de selección

La selección natural no sólo es oportunista y miope, sino también, con dos excepciones que pronto examinaré, ferozmente *individualista*. No favorece la especie ni la población, sino al organismo individual. Si una propiedad surgida de una mutación incrementa la aptitud relativa del organismo, quedará fijada en la población aun cuando también provoque una disminución de la aptitud absoluta. Imaginemos una población de peces, expuestos a depredadores, que inicialmente nadan en cardúmenes dispersos. Si una mutación hace que el pez en el que ocurre se desplace hacia el centro del cardumen, este individuo será menos vulnerable a la depredación y como consecuencia tenderá a dejar más descendencia. A medida que este comportamiento se difunde por la población, el cardumen se vuelve más compacto y resulta, así, un blanco más fácil para los depredadores. En cada etapa del proceso, es mejor buscar el medio que los bordes del cardumen. Sin embargo, en términos de aptitud absoluta el

resultado para todos es peor que la situación inicial, mientras que en términos de aptitud relativa no hay cambios. De manera similar, la selección sexual descontrolada es una explicación plausible de las grandes y disfuncionales astas encontradas en algunas especies de ciervos.

Una excepción al individualismo es la *selección por parentesco* (una forma de «subindividualismo»), en la cual la unidad de selección es el gen y no el organismo individual. La elección de la unidad no importa cuando el efecto de un gen incrementa simultáneamente y en la misma proporción la presencia de éste en la población y el número de descendientes del organismo que exhibe ese comportamiento. Así sucede, por ejemplo, en la evolución de una búsqueda más eficiente de alimentos. Pero en algunos casos el gen puede ser beneficioso aun cuando no lo sea en el organismo en el cual desencadena el comportamiento, esto es, cuando un organismo se «sacrifica» en bien de parientes cercanos que probablemente tengan el mismo gen. Cuando un animal observa a un depredador y emite una señal de alarma, sus posibilidades de supervivencia a menudo disminuyen, a la vez que aumentan las de los parientes cercanos que están en las proximidades. Como estos parientes, o algunos de ellos, también tendrán el «gen de la advertencia», sus mayores posibilidades de supervivencia quizás hagan que el gen se difunda en la población si ellos sobrecompensan las menores posibilidades del animal que ha emitido la señal. Ésta no es la única explicación posible de esa difundida forma de comportamiento, aunque en algunos casos es la más verosímil. En otros, las señales de alarma sirven para distraer a los depredadores o alertar a todos los coespecíficos (parientes o no). Y en un tercer grupo de casos, tal vez sirvan para *engañar* a los coespecíficos, por ejemplo, para inducirlos a alejarse de algún recurso escaso. En consecuencia, las señales de alarma no tienen nada de intrínsecamente altruista o autosacrificial, si bien pueden, en ciertas circunstancias, «remedar» el auténtico altruismo.

Otra excepción es la *selección grupal* (una forma de «supraindividualismo»). Consideremos dos poblaciones de peces, una en la cual la mutación centrípeta ha ocurrido y otra en la que no se ha producido. Con el tiempo, la primera tendrá menos descendencia que la segunda y, a la larga, podría ser desplazada. La selección parecería actuar en el nivel del grupo y no del individuo. No obstante, si las dos poblaciones coexisten, la segunda será vulnerable a invasores de la primera. Sea el comportamiento centrípeto causado por mutación o por inmigración, el resultado es el

mismo, a saber, el desplazamiento de los ejemplares que no se comportan de ese modo. De manera análoga, si los organismos de una población tienen un gen que les impide pastar en exceso, y evitan así la «tragedia de los pastos comunes», podrían ser superados en capacidad reproductiva por organismos menos inhibidos que carecen del gen. Por esa razón, no se ha considerado que la selección grupal sea un mecanismo plausible para generar cooperación o continencia. Sin embargo, a la luz de la teoría del castigo altruista expuesta en el Capítulo 15, esta objeción puede refutarse. Si los organismos de una población tienen un gen que los lleva a castigar a los que no cooperan, éstos no obtendrán ninguna ventaja reproductiva con su actitud «aprovechada». Con todo, este mecanismo sólo puede asegurar la cooperación donde es posible identificar con certeza a los individuos que no cooperan. Parece poco probable que sea válido para los peces con comportamiento centrípeto, pero bien podría aplicarse a animales que se niegan a compartir la comida con otros.

La selección por parentesco y la selección grupal proponen dos mecanismos para el surgimiento de un comportamiento cooperativo, el primero basado en los genes compartidos y el segundo en el castigo altruista. Un tercer mecanismo es el del *altruismo recíproco* o de «hoy por ti, mañana por mí» en interacciones reiteradas, como «te rasco la espalda; ráscame tú la mía» (expresión que debe tomarse al pie de la letra entre algunos animales) o «te ofrezco comida cuando me sobra; ofrécemela cuando tengas». La otra cara de la moneda es el castigo o, al menos, la abstención de la cooperación, cuando la otra parte omite actuar a la recíproca. Para que este mecanismo funcione, los individuos deben interactuar con la frecuencia suficiente para hacer valiosa la continencia, recordar lo que otros hicieron en anteriores ocasiones y reconocerlos cuando vuelven a encontrarlos.

Este mecanismo parece explicar la cooperación en muchos casos de interacción diádica; siguen algunos ejemplos y un contraejemplo. Cuando se brinda a dos monos capuchinos la oportunidad de compartir comida en ocasiones sucesivas, el segundo da más al primero cuanto más ha recibido de éste. En el caso de los chimpancés, es más probable que compartan la comida con quienes los han acicalado ese mismo día. Cuando dos estorninos se encuentran en una situación en la cual cada uno recibe comida sólo si y cuando el otro presiona una palanca, se alternan para hacerlo. Las hembras de un nido de vampiros regurgitan comida con sangre para otras

que no han logrado conseguirla en el pasado reciente. Los vampiros son capaces de reconocerse entre sí y es más probable que den sangre a aquellos que la han dado en el pasado, pero también a parientes. En otro experimento, cada una de dos urracas azules participaba de un programa que la recompensaba más cuando presionaba la «palanca egoísta» A que cuando presionaba la «palanca cooperativa» B, con prescindencia de lo que hiciera la otra, y ambas recibían más cuando las dos presionaban B que cuando ninguna lo hacía. Aquí no apareció una actitud del tipo «hoy por ti y mañana por mí»: tras alguna cooperación inicial, las urracas terminaron por presionar uniformemente la palanca A.

Sin embargo, el altruismo recíproco no es un mecanismo factible para generar cooperación en grupos más grandes. La única estrategia de castigo que podría funcionar es la de la respuesta del «gatillo» [*grim trigger*], por la cual la defección (conducta no cooperativa) de un miembro provoca de inmediato que todos los otros dejen de cooperar, tanto con el desertor como entre sí. Intuitivamente, esta reacción extrema no parece plausible, y en el plano empírico no parece ocurrir.

* * *

NOTA BIBLIOGRÁFICA

En «Selection by consequences», *Science*, 213(4507), 1981, págs. 501-504, Burrhus F. Skinner sostuvo la importancia de *tres* maneras de explicar la conducta por sus consecuencias: mediante la selección natural que actúa en los individuos, mediante el refuerzo y (aunque Skinner no utiliza esta expresión) mediante la selección grupal. Una eficaz introducción a la teoría del refuerzo es John E. R. Staddon, *Adaptive Behavior and Learning*, Cambridge y Nueva York, Cambridge University Press, 1983. En David Lee y Phillip Belfiore, «Enhancing classroom performance: a review of reinforcement schedules», *Journal of Behavioral Education*, 7(2), 1997, págs. 205-217, es un estudio del uso que se puede dar a la teoría del refuerzo para modelar (y no explicar) la conducta. Una exposición clásica de la teoría de la selección natural, notable por la insistencia en el carácter individualista de la selección, es George C. Williams, *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*, Princeton, Princeton University Press, 1966. En lo relacionado con el análisis del ascenso de gradientes y la «metáfora de los paisajes de aptitud»,

véase el Capítulo 2.4 de Sergey Gavrilets, *Fitness Landscapes and the Origin of Species*, Princeton, Princeton University Press, 2004. En Richard Dawkins, *The Selfish Gene*, 2ª edición, Oxford y Nueva York, Oxford University Press, 1990 [*El gen egoísta: las bases biológicas de nuestra conducta*, Barcelona, Salvat, 2002], se hallará una exposición que hace hincapié en el gen como unidad de selección. Una excelente introducción a las señales animales es William A. Searcy y Stephen Nowicki, *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems*, Princeton, Princeton University Press, 2005. Se encontrará un examen del posible papel del castigo altruista en la selección grupal en Ernst Fehr y Urs Fischbacher, «Social norms and human cooperation», *Trends in Cognitive Sciences*, 8(4), 2004, págs. 185-190. Un estudio pionero de la cooperación fundada en una actitud del tipo «hoy por ti, mañana por mí» entre animales no relacionados es Robert Axelrod y William Hamilton, «The evolution of cooperation», *Science*, 211(4489), 1981, págs. 1390-1396.

Notas al pie

[3] Recuérdese, sin embargo, que no siempre somos muy buenos para advertir cuál de dos experiencias fue más dolorosa (enigma 12 en el Capítulo 12).

[4] También es más fácil establecerlo cuando la tecnología del juego admite la posibilidad de casi aciertos. Aunque cada uno de éstos es menos reforzador que un verdadero acierto, en conjunto son más.

[5] Su buena suerte, en este caso, es su mala suerte, y la buena suerte del casino.

[6] ¿Mejores que qué? Nos ocuparemos de esta pregunta a la brevedad.

[7] La aptitud en sentido reproductivo difiere de la aptitud ecológica, tal como la mide la extensión de la vida. Aunque la aptitud ecológica suele ser un medio de la aptitud reproductiva, ambas también pueden divergir, como sucede cuando los padres corren un riesgo de muerte para defender su progenie.

[8] La radiación, por ejemplo, puede causar mutaciones; por esta razón, de suceder así, la exploración con rayos X para detectar un cáncer puede llegar a causar esta misma enfermedad.

[9] Como no he encontrado un ejemplo inglés equivalente, ilustro esta idea con dos frases noruegas: *«Vent, ikke heng ham»* y *«Vent ikke, heng ham»*, que significan, respectivamente: «Espera, no lo cuelgues» y «No esperes, cuélgalo».

[10] He escogido este ejemplo porque esos dos aminoácidos son los únicos codificados por un solo triplete. Todos los demás ácidos tienen dos o cuatro codones. Así, la frase «un camino del aminoácido X al aminoácido Y» debe leerse, a los presentes efectos, como «un camino de un codón determinado de X a uno de los codones de Y».

[11] Este tipo de juego lingüístico fue practicado por un grupo francés de escritores conocido como movimiento Oulipo. El más célebre de esos autores, Georges Perec, escribió toda una novela, *La disparition*, sin utilizar una sola vez la letra «e». [Hay traducción en castellano: *El secuestro*, Madrid, Luarna, 1969. En la versión castellana se ha eliminado la letra «a».]

Capítulo 17

Selección y comportamiento humano

Variación y selección

¿Cuánto del comportamiento humano es elegido o «seleccionado» por algún otro mecanismo que no sea la elección intencional? Para responder a esta pregunta, señalemos ante todo que cualquier mecanismo de selección necesita una materia prima, insumos, para trabajarla. Como el mismo proceso de selección, el origen de la variación puede ser intencional o no intencional (véase la Figura 17.1).

	Origen intencional de la variación	Origen no intencional de la variación
Selección intencional	Selección artificial en cultivo de plantas y crianza de animales	Mejora gradual de embarcaciones Eugenesia Aborto e infanticidio selectivos
Selección no intencional	Selección de empresas a través de la competencia en el mercado	Selección natural (Capítulo 16)

FIGURA 17.1

Variación no intencional, selección no intencional

La selección natural ha dado forma, sin duda alguna, a la estructura física de los seres humanos, que les brinda oportunidades para la acción y, a la vez, les impone limitaciones. Quienes tratan de explicar el comportamiento humano desde el punto de vista de la selección natural tienen a veces pretensiones más fuertes. Quieren explicar los patrones de conducta mismos, no simplemente las estructuras que los hacen posibles.

El mecanismo más plausible es el hecho de que la evolución haya producido *emociones* con sus tendencias características a la acción. Como un hombre nunca puede estar seguro de si es el padre de su progenie, mientras que la mujer no tiene ninguna duda de ser la madre, cabría esperar que la selección natural produjera una tendencia más fuerte a sentir celos sexuales en los hombres que en las mujeres. Muchas estadísticas de homicidios lo confirman. Así, de mil sesenta homicidios de cónyuges ocurridos en Canadá entre 1974 y 1983, ochocientos doce fueron cometidos por hombres y doscientos cuarenta y ocho por mujeres, pero entre los motivados por los celos hubo ciento noventa y cinco hombres responsables y apenas diecinueve mujeres. La teoría de la selección natural también predice que los padres mostrarán mayor compromiso emocional con sus hijos biológicos, portadores de sus genes, que con los hijos adoptivos. Los datos también confirman esta predicción. De tal modo, un niño estadounidense que viviera con uno o más progenitores sustitutos en 1976, tenía alrededor de cien veces más probabilidades de sufrir abusos que un niño que viviese con sus padres naturales. La selección natural puede propiciar asimismo la *falta* de emoción. En los humanos y otras especies de primates, los peligros de la endogamia se mantienen a raya por medio de la falta de atracción sexual entre los jóvenes que crecen juntos, estén o no relacionados entre sí.[12]

Al actuar sobre los grupos más que sobre los individuos, la selección natural puede haber favorecido también las emociones de la ira y la indignación hacia quienes violan las normas de la cooperación, motivando el castigo aun con algún coste para quien lo inflige (Capítulo 16). Una cuestión más enigmática es si la selección propició, y por qué, la emoción del *desprecio*, que se dirige hacia quienes violan normas sociales y no

morales. Como muchas normas sociales son arbitrarias y hasta disfuncionales (Capítulo 22), es difícil ver de qué modo podría sostenerlas la selección grupal. Dada una tendencia de los otros a excluir a quienes violan las normas sociales, el respeto de éstas sería más útil para la aptitud reproductiva si el coste del ostracismo superara el beneficio obtenido con la transgresión de la norma. El enigma es, ante todo, por qué habría de surgir esa tendencia. Por ejemplo, ¿por qué la gente desaprobaría el adulterio? Las normas sociales contra éste implican reacciones de terceros que difieren de la reacción de celos sexuales del segundo. Aunque A pudiera beneficiarse con la desaprobación de C de las insinuaciones de B al cónyuge de A, ese beneficio no crea una presión selectiva para que C se comporte de ese modo. Si bien la selección grupal podría favorecer los genes que inducen el castigo de los aprovechados por parte de terceros, es menos evidente el beneficio para el grupo del castigo aplicado por terceros al adulterio. Aunque la tendencia de las normas contra el adulterio femenino al ser más fuertes que las contrarias al adulterio masculino sugiera una explicación evolutiva, cuesta ver cuál sería el mecanismo.[13]

Otras tesis son más especulativas, como la idea de que, en los seres humanos, el *autoengaño* se desarrolló debido a sus beneficios evolutivos. Así reza el argumento. A menudo es útil engañar a otros. Sin embargo, el engaño deliberado o hipócrita es difícil de llevar a cabo. En consecuencia, el autoengaño evolucionó para facilitar a la gente engañar con éxito a otros. La debilidad del argumento radica en que si el autoengaño nos lleva a sostener falsas creencias, éstas podrían tener consecuencias desastrosas si se las utilizara como premisas del comportamiento. Nadie ha argumentado de manera convincente que el resultado *neto* de esos efectos opuestos tiende a ser positivo, como tendría que serlo para que el autoengaño realzara la aptitud evolutiva.

Aún más especulativa es la tesis de que la *depresión* unipolar puede haberse desarrollado como una herramienta de negociación, más o menos parecida a una huelga laboral. Por ejemplo, una presunta función de la depresión posparto consiste en inducir a otros a compartir la crianza del niño, así como los trabajadores van a la huelga para obligar a los empleadores a compartir las ganancias. Según este punto de vista, los suicidios inducidos por la depresión son el coste de plantear una amenaza creíble de suicidio. Se trata, por decirlo de algún modo, de intentos de suicidio que no logran fracasar. El insomnio se explica como una

asignación de recursos cognitivos para resolver la crisis a la que responde la depresión, mientras que el hipersomnio (dormir más de lo normal) se presenta como una manera de reducir la productividad y, de tal modo, fortalecer la eficacia negociadora de la depresión. El argumento, si bien compatible con algunos hechos conocidos acerca de esta última, ignora muchos otros, por ejemplo que tanto ella como el suicidio «se llevan en la sangre», que los individuos divorciados (sin pareja negociadora) son más propensos a deprimirse que los casados o quienes nunca se casaron, y que los sucesos estresantes de la vida no son una condición necesaria ni suficiente de la depresión.

La explicación de la depresión como una herramienta de negociación es otro ejemplo de la búsqueda ubicua de un *significado* o una *función* de todos los comportamientos aparentemente inútiles o disfuncionales. Hasta cierto punto, la búsqueda de un significado es una buena estrategia de investigación; superado ese punto, se convierte en artificiosa y, como en algunos de los ejemplos mencionados, definitivamente absurda. Los rasgos perjudiciales tienen tantas maneras de preservarse en una población, que no se puede dar por descontado que los comportamientos de aparición frecuente confieran aptitud reproductiva al agente.[14] La selección natural ha favorecido sin duda la propensión a sentir dolor físico, y no hay una razón a priori por la cual no pueda promover la tendencia a experimentar dolor mental. Pero para establecer la función de la depresión no basta con proponer un relato del tipo «así es porque así es» que dilucide algunos de los rasgos conocidos de la enfermedad. Lo crucial es que la hipótesis explique también hechos que vayan más allá de los que está destinada a explicar (Capítulo 1), y preferentemente «nuevos hechos» que eran desconocidos antes de que ella los predijera.

Variación intencional, selección intencional

Aunque este mecanismo doblemente intencional ha sido y seguirá siendo (no podemos sino esperarlo) poco importante para explicar los patrones del comportamiento *humano*, merece la pena, no obstante, considerar su estructura.

En *El origen de las especies*, Darwin escribió que la «naturaleza da variaciones sucesivas; el hombre las acumula en ciertas direcciones que le son útiles». Pero no se trata de un simple caso de «la naturaleza propone y

el hombre dispone», pues, como también señaló el propio Darwin, el comportamiento humano puede modificar el insumo:

> Un alto grado de variabilidad es evidentemente favorable, pues aporta en forma ilimitada los materiales para que la selección pueda elaborarse; y no es que las simples diferencias individuales no sean lo suficientemente amplias para permitir, con sumo cuidado, la acumulación de una gran cuantía de modificación en casi todas direcciones que se deseen. Pero como las variaciones manifiestamente útiles o agradables al hombre aparecen sólo de cuando en cuando, la probabilidad de su aparición aumentará mucho más cuando se tenga un gran número de individuos. Por consiguiente, el número es de suma importancia para el éxito. Según este principio, Marshall ha señalado, con respecto a las ovejas de ciertas zonas de Yorkshire, que «como en general pertenecen a los pobres, y están en su mayor parte en pequeñas parcelas, nunca pueden mejorar». Por otro lado, quienes trabajan en viveros, al cultivar grandes cantidades de las mismas plantas, tienen generalmente mucho más éxito que los aficionados en la obtención de nuevas y valiosas variedades.

Hoy podemos agregar que también es posible mejorar la selección artificial mediante la inducción de mutaciones. Por otra parte, el mantenimiento de «bibliotecas genéticas» puede impedir la reducción de la variedad genética que, de lo contrario, es el resultado inevitable de la selección de rasgos particulares.

Con referencia al proceso mismo de la selección, Darwin distinguía entre dos *niveles de intencionalidad*:

> Hoy en día, los más eminentes criadores tratan por medio de una selección metódica, y con un objeto determinado, de formar una nueva subcasta o estirpe superior a cuantos géneros existan en su país. Pero para nuestro propósito, es más importante una forma de selección que podría llamarse inconsciente, y que resulta de todo el que intenta poseer los mejores animales y hacerlos reproducir. Así, el hombre que quiere tener perros de muestra, naturalmente trata de elegir perros buenos y después criar con los mejores, aunque no tenga el deseo ni la esperanza de modificar permanentemente la casta.

Variación no intencional, selección intencional

Hay muchos casos en los que un nuevo organismo o una nueva forma surge por accidente y luego es aceptado o rechazado sobre la base de la elección intencional. Mientras que la selección natural tiende a producir un número igual de organismos masculinos y femeninos, el infanticidio en función del género y, en épocas más recientes, el aborto sobre esa misma base, pueden generar un grave desequilibrio de los sexos en la población. Sólo en la India y China «faltan» alrededor de ochenta millones de mujeres por esa razón. Las políticas eugenésicas se han utilizado en vasta escala con

el fin de impedir la reproducción de los enfermos y retrasados mentales. En la Alemania nazi, entre trescientos mil y cuatrocientos mil individuos fueron esterilizados a la fuerza por ese motivo. Con el perfeccionamiento de las técnicas de exploración prenatal, el aborto selectivo puede llegar a ser un determinante importante en la composición de las poblaciones humanas. Si nuevos progresos hacen posible determinar el sexo del niño en el momento de la concepción, la elección intencional terminará por reemplazar la selección.

La variación aleatoria, combinada con la selección intencional, también puede dar forma al desarrollo de artefactos. Cuando el ministro y sociólogo noruego Eilert Sundt visitó Inglaterra en 1862, tomó conocimiento de la teoría de la selección natural de Darwin (publicada en 1859) y se propuso aplicar una variante a la construcción naval:

> Un constructor de barcos puede ser muy diestro y, pese a ello, no lograr construir jamás dos embarcaciones exactamente iguales, aun cuando se aplique a ese fin. Las variaciones surgidas de esta manera pueden calificarse de *accidentales*. Pero incluso una variación muy pequeña suele advertirse durante la navegación, y entonces *no es accidental* que los marinos lleguen a *reparar en* esa embarcación que ha mejorado o es más conveniente para sus finalidades, y que aconsejen *elegirla* como modelo por *imitar*. […] Tal vez creamos que cada una de esas embarcaciones es perfecta a su manera, dado que ha alcanzado la perfección en virtud de un desarrollo unilateral en una dirección específica. Todos los tipos de mejora han progresado al extremo de hacer que nuevas transformaciones impliquen defectos capaces de ser algo más que contrapesos a sus ventajas. […] Y concibo el proceso del siguiente modo: una vez nacida la idea de nuevas y mejoradas formas, *una larga serie de experimentos prudentes*, cada uno de los cuales entraña cambios sumamente pequeños, puede conducir al afortunado resultado de que, del cobertizo del constructor, surja una embarcación cuyo igual todos deseen.

En este texto, Sundt mejoró la idea de Darwin en un aspecto crucial.[15] En tanto que Darwin confesaba su ignorancia sobre el origen de la variación, el ministro noruego dio con la idea de situar su fuente en los *errores de replicación*, similares a los errores tipográficos y a las (que hoy conocemos como) mutaciones en el ADN. La imperfección del constructor naval (su incapacidad para hacer copias perfectas) es una condición de la perfección última del resultado final. Sundt tiene la precaución de señalar que el resultado del proceso es un máximo local, a partir del cual no pueden surgir nuevas mejoras por obra de la acumulación de cambios. En la última frase, también sugiere que el proceso puede convertirse en una selección artificial, cuando la gente se consagra a hacer experimentos deliberados en vez de permitir que las variaciones se produzcan por azar. Al igual que Darwin, Sundt señaló que la inteligencia o la intencionalidad pueden

aparecer en dos niveles: en primer lugar, cuando la gente *advierte* que un modelo es más marinero que otro anterior, y luego cuando *entiende* que las mejoras podrían acelerarse si la variación al azar fuera reemplazada por experimentos sistemáticos.[16]

Variación intencional, selección no intencional

El funcionamiento de los mercados económicos tiene algunas características en común con la selección natural. La analogía tiene dos versiones, una relativamente cercana a esta selección, y otra más alejada. Ambas comparten la premisa de que, vistas las numerosas limitaciones de la racionalidad humana, las empresas o los administradores son *ineficientes*, por su incapacidad de estimar cuáles serán las decisiones de producción y comercialización que maximizarán su ganancia. Pese a ello, el mecanismo del mercado erradicará las empresas ineficientes, de modo que en cualquier momento dado se observará esencialmente la presencia de empresas eficientes. Todo sucede «como si» los administradores fueran eficientes.

En la primera y más simple de las versiones, todas las empresas *tratan* sin descanso de aumentar sus ganancias mediante procesos de imitación e innovación. Aunque por sí misma la imitación no genera nuevos insumos para que la selección actúe sobre ellos, la imitación *imperfecta*, como se ha señalado, puede tener este resultado. La innovación también es, por definición, una fuente de nuevos insumos. Cuando, gracias a la suerte lisa y llana, la innovación o la imitación imperfecta permiten a una empresa producir a menor coste, ésta puede vender a precios más bajos que sus rivales y expulsarlos de la actividad a menos que también ellos adopten los métodos más eficientes. A través de uno u otro mecanismo, la quiebra o la imitación, esas técnicas eficientes se difundirán por la población de empresas. Si suponemos que tanto la imitación como la innovación ocurren preponderantemente por pequeños pasos y que la competencia tiene lugar en un entorno que en otros aspectos es constante, la consecuencia será un máximo local de ganancias de equilibrio.

La segunda versión niega que las empresas intenten siempre maximizar las ganancias. Antes bien, las empresas apelan a *rutinas* o reglas prácticas que se mantienen mientras las ganancias se sitúan en un nivel «satisfactorio». Para decirlo con un neologismo, «satisfizan» [*«satisfice»*] en vez de maximizar. El significado de esta actitud depende quizá de

muchos factores, pero para simplificar podemos suponer que una empresa cuyas ganancias están de manera uniforme por debajo del nivel satisfactorio irá a la quiebra o enfrentará la amenaza de una adquisición hostil. La rutina más simple consiste en hacer todo como antes siempre que las ganancias se mantengan en un nivel «satisfactorio». Rutinas más complicadas podrían implicar la fijación de precios mediante un margen de beneficios constante sobre los costes o la inversión de un porcentaje determinado de las ganancias en nuevos productos. La idea de la «satisfización» se refleja en dichos como «equipo que gana no se toca» o «si no está roto, no lo arregle». Desde cierto punto de vista, la satisfización podría incluso ser óptima. Según una frase que he citado antes, «la mayor de todas las ganancias monopólicas es una vida tranquila».

Supongamos ahora que las ganancias caen por debajo del nivel satisfactorio. Una empresa que ha hecho lo mismo año tras año puede ser víctima de un análogo organizacional de la herrumbre o la esclerosis. Las conmociones externas, como un alza en los precios del petróleo o la modificación de una tasa de cambio importante, pueden incrementar los costes o reducir los ingresos. Es posible que la demanda de los consumidores cambie; los competidores quizá den con mejores métodos o nuevos productos, o los trabajadores pueden someter a la empresa a una costosa huelga. Sea cual fuere la causa (que puede ser incluso desconocida para la firma en cuestión), las ganancias insatisfactorias promoverán la búsqueda de nuevas rutinas mediante alguna combinación de innovación e imitación. Es probable que uno y otro procedimiento sean primordialmente locales, en el sentido de limitarse a alternativas cercanas a las rutinas existentes. Los grandes cambios, cualquiera sea su tipo, pueden ser demasiado costosos para una empresa que atraviesa dificultades financieras (Capítulo 9), y las innovaciones no graduales también son más exigentes desde un punto de vista conceptual.

Como es obvio, el proceso de imitación tiende a remitirse a los competidores exitosos. Que la innovación sea aleatoria o dirigida dependerá de las causas percibidas de la crisis que la ha motivado. Si la caída de las ganancias por debajo del nivel aceptable es la resultante de un alza en los precios del petróleo, es posible que la empresa encauce su búsqueda hacia métodos de ahorro de combustible.[17] Si es la consecuencia de una modificación de la tasa de cambio entre el dólar y el euro, es más probable que la empresa actúe al azar. En todos los casos, sin embargo, su

comportamiento tendrá un vigoroso componente intencional. La decisión de modificar las rutinas en vigencia es intencional, como lo es la determinación de la suma que se invertirá en innovación o imitación. La elección de modelos por imitar es deliberada y, como recién se ha señalado, la empresa puede orientar intencionalmente la búsqueda de nuevas rutinas en una dirección específica.

Las nuevas rutinas resultantes de este proceso quedan luego expuestas a las fuerzas ciegas de la competencia en el mercado. Si permiten a la empresa alcanzar un nivel satisfactorio de ganancias, la búsqueda se cancelará hasta que estalle una nueva crisis. De no llegar a ese nivel satisfactorio, la firma puede volver a intentarlo o verse obligada a declarar la quiebra. Tarde o temprano, las empresas que no «satisfizan» son eliminadas. De por sí, este proceso no tiende a producir empresas maximizadoras de las ganancias. Para ver cómo podría suceder esto, debemos incorporar más explícitamente la *competencia* al cuadro. Si suponemos que, como parte de sus rutinas, las firmas invierten un porcentaje fijo de las ganancias en nuevos productos, aquellas que por pura suerte hayan encontrado una rutina más adecuada que sus competidores se expandirán de tal manera que, con el tiempo, sus rutinas tendrán una presencia más abundante en la población de empresas.[18]

Modelos de selección en las ciencias sociales

La utilidad de estos modelos depende de una sencilla pregunta empírica: ¿cuál es la velocidad con que se elimina a las empresas ineficientes, comparada con la velocidad de cambio del entorno? En el capítulo anterior planteé el mismo interrogante con respecto a la selección natural y presenté el argumento indirecto de que la adaptación muy fina de los organismos a su medio ambiente sugiere que éste debe haber cambiado de manera relativamente lenta. En el caso del medio ambiente económico, podemos hacer una evaluación más directa. En el mundo moderno, las empresas están expuestas a cambios de una velocidad sin precedentes. Si quedaran reducidas a hacer una exploración gradual del entorno, serían crónicamente inadecuadas. Las firmas exitosas son, más probablemente, las capaces de apuntar por delante del blanco y, así, *prever* el cambio. También esta estrategia fracasará gran parte de las veces, pero al menos no lo hará siempre. Por otra parte, debido a su influencia política, las grandes

corporaciones también pueden tener la aptitud de *dar forma* al medio ambiente en el cual operan. En una etapa anterior del capitalismo despiadado entre pequeñas empresas, los mecanismos de selección del tipo que he descrito tal vez hayan sido o no importantes: no lo sabemos. En nuestros días, es improbable que expliquen gran parte de lo que observamos.

Hay asimismo en juego un problema más general. Cuando se los ataca por la falta de realismo de sus supuestos, los teóricos de la elección racional responden de manera habitual que su única aspiración es explicar el comportamiento bajo el supuesto de que la gente actúa «como si» maximizara la utilidad (o la ganancia, o cualquier otra meta). A menudo, agregan que este supuesto puede justificarse en virtud de algún tipo de mecanismo de selección. En la esfera económica, se presume que la tarea queda a cargo de la competencia en el mercado. He sostenido que no puede ser así.[19] La manera más general de enunciar mi objeción es tal vez decir que, aun cuando pudiera mostrarse que la competencia en el mercado mejora la eficiencia a través de la eliminación de las empresas ineficientes, hay un gran paso desde «mejorar la eficiencia» hasta la ultrasofisticada maximización del tipo «como si» atribuida a las firmas en los modelos económicos.

En la esfera política, se atribuye a la competencia electoral la capacidad de garantizar que los únicos políticos a quienes observamos sean los elegidos o reelegidos; uno puede suponer, por ejemplo, que todos los políticos actúan «como si» sólo les preocuparan sus perspectivas electorales. Sin embargo, el salto de una preocupación por las elecciones a un interés *exclusivo* en ellas no se justifica. Un vistazo sin prejuicios metodológicos a la política sugiere que hay tres clases de actores políticos: los oportunistas (a quienes sólo interesa ser elegidos), los reformadores (que se interesan en la implementación de sus políticas) y los activistas o militantes (cuya mayor preocupación es «dejar sentada una posición»).[20] La concepción según la cual la política se basa en la interacción entre estos tres grupos dentro de cada partido, y entre distintos partidos, es claramente más realista que el punto de vista que la concibe de acuerdo con el modelo del «puesto de helados» (Capítulo 19), para el cual los partidos resueltos a maximizar los votos convergen hacia el centro. Se encontrará una sorprendente refutación de la tesis de que los políticos sólo están motivados por preocupaciones electorales en la estirpe de políticos franceses iniciada

por Jean Jaurès y continuada por Léon Blum, Pierre Mendès-France y Michel Rocard, todos los cuales tenían como motivación transparente el deseo de promover los valores imparciales de la justicia social y la eficiencia económica. Es preciso decir, con todo, que en el caso de Rocard su disgusto por la política electoral conspiró contra su eficacia política.

Al margen de las arenas de la competencia, la racionalidad del «como si» tiene aún menos justificación. Las decisiones de consumo, el comportamiento electoral, la concurrencia a la iglesia, la elección de carrera y la mayoría de los demás comportamientos que podríamos nombrar no están sujetos a mecanismos de selección que remedan la racionalidad. Están sometidos, por supuesto, a *restricciones* que pueden reducir la importancia de la elección en general y de la elección racional en particular (Capítulo 9). Las restricciones actúan a priori, para hacer que ciertas elecciones sean inviables. La selección actúa a posteriori, para eliminar a quienes han hecho ciertas elecciones. Aunque ambos mecanismos contribuyen a la explicación del comportamiento, no pueden, ni en conjunto ni por separado, dar razón de éste en su totalidad. La elección sigue siendo el concepto medular de las ciencias sociales.

* * *

NOTA BIBLIOGRÁFICA

Los datos sobre las estadísticas de homicidios y el abuso infantil provienen de Martin Daly y Margo Wilson, *Homicide*, Nueva York, Aldine de Gruyter, 1988 [*Homicidio*, México, Fondo de Cultura Económica, 2003]. Se encontrarán objeciones a su explicación en David J. Buller, *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*, Cambridge (Massachusetts), The MIT Press, 2005. Con respecto a dos aspectos del argumento del autoengaño, véase Robert Trivers, *Social Evolution*, Menlo Park (California), Benjamin-Cummings, 1985 (favorable a una explicación evolutiva), y Vangipuram S. Ramachandran y Sandra Blakeslee, *Phantoms in the Brain: Human Nature and the Architecture of the Mind*, Nueva York, Quill, 1998 [*Fantasmas en el cerebro: los misterios de la mente al descubierto*, Madrid, Debate, 1999] (opuesto a esa explicación). En lo que concierne a dos aspectos de la naturaleza adaptativa de la depresión, véase Edward H. Haggen, «The bargaining model of depression», en Peter Hammerstein (comp.), *Genetic and Cultural*

Evolution of Cooperation, Cambridge (Massachusetts), The MIT Press in cooperation with Dahlen University Press, 2003, págs. 95-123 (favorable a una explicación evolutiva), y Peter D. Kramer, *Against Depression*, Nueva York, Viking, 2005 [*Contra la depresión*, Barcelona, Seix Barral, 2006] (opuesto a esa explicación). El análisis de los mercados en función de selección natural tiene su origen en Armen Alchian, «Uncertainty, evolution, and economic theory», *Journal of Political Economy*, 58(3), junio de 1950, págs. 211-221. Su versión más sofisticada (que *no* respalda la maximización del tipo «como si») es Richard Nelson y Sidney Winter, *An Evolutionary Theory of Economic Change*, Cambridge (Massachusetts), Harvard University Press, 1982. La teoría de la «satisfización» proviene de Herbert Simon, «A behavioral theory of rational choice», *Quarterly Journal of Economics*, 69, 1954, págs. 99-118. La economía de los deportes de equipo es el tema de David J. Berri, Martin B. Schmidt y Stacey L. Brook, *The Wages of Wins: Taking Measure of the Many Myths in Modern Sport*, Stanford, Stanford University Press, 2006. La distinción entre oportunistas, reformadores y activistas proviene de John E. Roemer, *Political Competition: Theory and Applications*, Cambridge (Massachusetts), Harvard University Press, 2001. Sobre los argumentos «como si», véase también la nota bibliográfica del Capítulo 1.

Notas al pie

[12] En consecuencia, el tabú del incesto quizá se refiera a una tentación que existe con menos frecuencia de lo que se ha creído. Freud, en contraste, suponía que ese tabú había surgido para contrarrestar un deseo inconsciente de mantener relaciones sexuales con parientes cercanos.

[13] Pero cualquier sociobiólogo merecedor del pan que come probablemente podría dar con una explicación en menos de media hora.

[14] Además de los comportamientos menos que óptimos causados por los retrasos y la pleiotropía (Capítulo 16), esos rasgos pueden ser mantenidos por diversos mecanismos genéticos relacionados con la presencia en los organismos que se reproducen sexualmente de dos variantes diferentes (alelos) de cada gen.

[15] La mejora sólo fue posible, desde luego, porque Sundt abordaba un problema diferente, dado que en 1862 nadie tenía los medios conceptuales necesarios para imaginar que la fuente de la variación en los *organismos* podía ser la replicación aleatoria de errores. Este salto sólo resultó posible luego de que Mendel revelara la naturaleza discreta de las unidades de la herencia (los genes) y de que Watson y Crick demostraran que la replicación intervenía en el proceso de la herencia. Me pregunto qué habría respondido Darwin si Sundt le hubiese preguntado si la imperfección de la maquinaria reproductiva no podía ser el origen de la variación biológica.

[16] De esta manera, también se podría prevenir la desventurada situación que surgiría si los constructores navales llegaran a ser tan buenos que nunca cometieran errores.

[17] A menos que se esperen nuevos incrementos, la racionalidad no exige que la empresa busque innovaciones *que economicen combustible*, dado que nadie puede saber cómo es el conjunto de innovaciones factibles (véase también el Capítulo 14). No obstante, el aumento de los precios del petróleo tenderá a destacar la importancia de esas innovaciones.

[18] En esta versión, no procurarán desalojar en forma deliberada a sus rivales del mercado utilizando, por ejemplo, las elevadas ganancias para vender por debajo del coste hasta que los otros se rindan, dado que no están interesadas en obtener ganancias más que satisfactorias.

[19] Por añadidura a los argumentos generales que he mencionado, la economía de los deportes de equipo plantea una posible objeción a la idea de que la selección provoca la maximización de las ganancias. Si los equipos de béisbol o fútbol maximizadores de la ganancia utilizaran sus utilidades para comprar a todos los mejores jugadores de la liga en que participan, su superioridad sería tan abrumadora que los partidos perderían gran parte de su incertidumbre y por lo tanto de su diversión, y con ello, de su aptitud de generar ganancias.

[20] En términos más formales, los tres grupos pueden distinguirse de la siguiente manera. Los oportunistas prefieren proponer la política A a la política B cuando la probabilidad de ganar con A es más grande que la probabilidad de ganar con B, dado que el partido de oposición propone una C fija. Los militantes prefieren proponer A a B cuando el afiliado partidario común y corriente extrae una utilidad más alta de A que de B (con prescindencia de lo que sea C). Los reformistas prefieren proponer A a B, habida cuenta de que el partido opositor propone C, cuando la utilidad esperada del afiliado común y corriente es más elevada en A que en B. De tal modo, a los oportunistas sólo les interesan las probabilidades; a los activistas, sólo las utilidades, y a los reformistas, unas y otras.

Quinta parte

Interacción

La interacción social puede adoptar muchas formas. (1) El resultado, para cada agente, depende de los resultados para otros. Esta interdependencia de los resultados puede manifestarse si el bienestar material o psíquico de otros afecta mi propio bienestar psíquico (Capítulo 5). (2) El resultado de cada uno puede depender de las acciones de todos. Esta interdependencia refleja una causalidad social general (Capítulo 18), ilustrada en fenómenos tales como el calentamiento global (provocado por el hombre). (3) La acción de cada uno depende de las acciones (previstas) de todos. Esta interdependencia es el tema específico de la *teoría de los juegos* (Capítulo 19 y Capítulo 20), que también integra a su marco los incisos 1) y 2). (4) Las creencias de cada uno dependen de las acciones de todos. Esta interdependencia puede surgir por una diversidad de mecanismos, como la «ignorancia pluralista» o las «cascadas informacionales» (Capítulo 23). (5) Las preferencias de cada uno dependen de las acciones de todos. Esta interdependencia es tal vez el aspecto peor entendido de la interacción social. Si bien abordo algunos aspectos de la cuestión en varios lugares, sobre todo en el Capítulo 22, no propongo una explicación generalizada.

Estas interdependencias pueden aparecer por medio de la acción descentralizada de individuos que no guardan una relación organizada entre sí (Capítulo 24). Sin embargo, gran parte de la vida social está más estructurada. Muchos resultados se producen en virtud de procedimientos de toma colectiva de decisiones, la discusión, la votación y la negociación, por medio de los cuales grupos de individuos llegan a decisiones que son vinculantes para todos ellos (Capítulo 25). Por último, las organizaciones actúan en función de reglas destinadas a armonizar los incentivos a los individuos y las metas del sistema (Capítulo 26).

Capítulo 18
Consecuencias indeliberadas

Consecuencias indeliberadas del comportamiento individual

Las cosas no siempre resultan como pretendemos. Muchos sucesos ocurren de manera no intencional. A veces, las causas son triviales, como cuando apretamos el acelerador en lugar del freno o pulsamos por error la tecla de «suprimir». Algunos mecanismos, no obstante, son más sistemáticos. Si bien es muy difícil que haya una «teoría general de las consecuencias indeliberadas», podemos al menos empezar a compilar un catálogo. Considero casos en los cuales las consecuencias no sólo son indeliberadas, sino también imprevistas. Los «efectos secundarios previsibles de la acción» no se buscan por sí mismos, sobre todo si son negativos, pero no los incluiré entre las «consecuencias indeliberadas de la acción».

Esas consecuencias indeliberadas pueden surgir tanto del comportamiento individual como de la interacción social. Comenzando por el primero, podemos utilizar una simple ampliación del marco de los deseos y las oportunidades que hemos expuesto en el Capítulo 9 (véase la Figura 18.1).

Si bien las acciones son configuradas por los deseos (o las preferencias), también ellas pueden configurar estos últimos. Así, además de la consecuencia deliberada de una acción, a veces hay una consecuencia indeliberada: un cambio de deseo. La adicción es un buen ejemplo. Bajo la influencia de drogas adictivas, la gente comienza a descontar más

ampliamente el futuro, y debilita de ese modo el efecto disuasivo del perjuicio provocado a largo plazo por la adicción. De haber sido previsto, este efecto podría haber impedido al agente internarse en el camino de la adicción, pero lo común es que no se prevea. Fenómenos similares se observan en situaciones más corrientes. Voy a la fiesta con la intención de tomar sólo dos tragos para poder conducir el auto de regreso a casa, pero luego de la segunda copa mi resolución se disuelve en el alcohol y bebo una tercera. De haberlo sabido, podría haber bebido una sola.

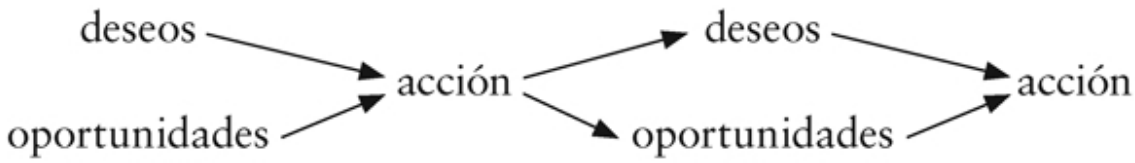


FIGURA 18.1

Los deseos también pueden ser afectados por una preferencia inconsciente por la novedad o el cambio («la hierba es más verde del otro lado del cerco»). En el cuento de Hans Christian Andersen «Lo que hace el padre bien hecho está», el granjero va al mercado a la mañana a vender o intercambiar su caballo. Al principio se encuentra con un hombre que tiene una vaca, y ésta le gusta tanto que, para hacerla suya, entrega el caballo a cambio. En transacciones sucesivas, cambia la vaca por una oveja, la oveja por un ganso, el ganso por una gallina y, por último, la gallina por un saco de manzanas podridas. Su camino a la ruina está pavimentado de mejoras escalonadas. En cada ocasión el granjero cree quedar en mejor situación gracias al intercambio, pero el resultado neto de todos ellos es desastroso.[1]

El «efecto de dotación», una implicación de la aversión a la pérdida (Capítulo 12), también ilustra el cambio de preferencias inducido por la elección pero indeliberado.[2] Muchos bienes adquieren para el propietario un valor subjetivo más grande que el que tenían antes de comprarlos, como lo muestra el hecho de que su precio mínimo de venta suele superar su precio máximo de compra por un factor que oscila entre dos y cuatro. Habida cuenta de que la mayoría de los bienes se evalúan como una pérdida cuando se renuncia a ellos y como una ganancia cuando se adquieren, y visto que las pérdidas cuentan más que las ganancias de la misma magnitud, eso es lo

que la aversión a la pérdida pronostica. Además, los experimentos muestran que los compradores en perspectiva subestiman el precio mínimo de reventa que aceptarían, una demostración de que el cambio de preferencias es efectivamente imprevisto.[3] Otro mecanismo que podría producir este «efecto de refuerzo», a saber, la tendencia a ver nuestras elecciones bajo una luz positiva una vez que las hemos hecho, es propuesto por la teoría de la disonancia cognitiva (Capítulo 1).

Para mostrar cómo puede la acción dar forma a las *oportunidades* de un modo indeliberado e imprevisto, consideremos el caso del pendenciero que es capaz de salirse con la suya en las transacciones con otros porque éstos, por lo común, prefieren ceder a hacerle frente. Tal vez, como consecuencia indeliberada de su conducta, los otros lo eviten, de modo que tendrá menos oportunidades de transacción con ellos. Cada encuentro es para él un éxito, pero cada vez tiene menos encuentros. Esta última consecuencia puede ser no sólo indeliberada e imprevista, sino también inadvertida. A juicio del pendenciero, el matonismo da resultado.[4] Aunque advierta los efectos negativos de su comportamiento, quizá persevere en él si los efectos positivos son mayores. De ser así, las consecuencias negativas serán previstas pero no buscadas por sí mismas.

Con frecuencia, la elección de una opción hoy elimina ciertas opciones del conjunto factible mañana. Este efecto puede ser previsto: mi restricción presupuestaria tal vez me permita comprar un automóvil, pero no dos. En ocasiones, sin embargo, es posible que el agente no sepa que la elección tiene consecuencias irreversibles. Un campesino puede tener tierras en las que hay algunos árboles y algunos campos. Con el fin de tener más terrenos de cultivo y madera para quemar, tala los árboles. La deforestación genera erosión, y de resultas el campesino se queda con menos tierra para cultivo que al comienzo. En una serie de ejemplos que examinaré a la brevedad, la erosión puede ser producto de un comportamiento *colectivo*, en caso de ocurrir, digamos, en los terrenos del granjero si y sólo si tanto él como sus dos vecinos causan una deforestación. Pero también es posible y bastante corriente que un individuo socave por sí solo y sin saberlo sus oportunidades futuras de acción. La culpa recae en un *déficit cognitivo*: el agente no puede predecir las consecuencias futuras del comportamiento presente. En otros casos, la culpa recae en un *déficit motivacional*: el agente atribuye escaso peso a las consecuencias futuras (conocidas y seguras) en comparación con las ganancias inmediatas (Capítulo 6).

Externalidades

Querría ahora ocuparme de las consecuencias indeliberadas de la *interacción*, un tema que fue una de las ideas clave en los comienzos de las ciencias sociales. Según la memorable frase de Adam Ferguson, la historia es «el resultado de la acción humana, pero no la ejecución del designio humano». Su contemporáneo Adam Smith se refirió a una «mano invisible» que da forma a los asuntos humanos. Medio siglo después, Hegel invocó la «astucia de la razón» para explicar el progreso de la libertad en la historia. Más o menos en la misma época, Tocqueville planteó, de manera similar, la idea de que en el progreso de la democracia, «todos los hombres la han ayudado con su esfuerzo: los que tenían el proyecto de colaborar para su advenimiento y los que no pensaban servirla; los que combatían por ella, y aun aquellos que se declaraban sus enemigos». Unos años más adelante, Marx aludió a la «alienación» del pueblo con respecto a su propia acción, y afirmó que «esta plasmación de las actividades sociales, esta consolidación de nuestros propios productos en un poder material erigido sobre nosotros, sustraído a nuestro control, que levanta una barrera ante nuestra expectativa y destruye nuestros cálculos, es uno de los momentos fundamentales que se destacan en todo el desarrollo histórico anterior».

Entre estos autores, sólo Adam Smith y Marx presentaron mecanismos específicos para la producción de consecuencias indeliberadas. En un lenguaje moderno, destacaron que las *externalidades* del comportamiento pueden agregarse para generar resultados no deliberados ni previstos por los agentes. En una forma estilizada, imaginemos que cada uno de muchos agentes idénticos emprende cierta acción para promover sus intereses. Como subproducto de esa acción, también impone un pequeño coste u otorga un pequeño beneficio (una externalidad negativa o positiva) a cada uno de los restantes (y a sí mismo). Cada agente, entonces, es el blanco de muchas de esas acciones. Mediante la adición de los efectos y, luego, la adición de la suma al beneficio privado del agente causado por su acción, llegamos al resultado final que los agentes generan por intermedio de sus acciones. Como suponemos que son idénticos, sus estados iniciales, los estados que cada uno de ellos pretende provocar y los estados que producen colectivamente pueden representarse por un único número, x, y y z, respectivamente.[5]

Supongamos en primer lugar que $z > y > x$, una externalidad positiva. Ése era el principal interés de Adam Smith: cuando un agente dirige su

> industria de tal forma que su producto represente el mayor valor posible, sólo piensa en su propia ganancia; pero en éste como en otros muchos casos, es conducido por una mano invisible a promover un fin que no entraba en sus intenciones. Mas no implica mal alguno para la sociedad que tal fin no forme parte de sus propósitos, pues al perseguir su propio interés, él promueve el de la sociedad de una manera más efectiva que si esto entrara en sus designios.

En la competencia en el mercado, el objetivo de cada empresa es obtener una ganancia gracias a producir más barato que sus rivales, pero al actuar de ese modo también se benefician los clientes. Éstos, en su carácter de trabajadores o administradores, podrían asimismo encontrarse en condiciones similares de beneficiar a otros a través de sus esfuerzos competitivos. El resultado ha sido un espectacular crecimiento secular. Previsto o imprevisto, es indudable que el efecto «no formaba parte» de sus intenciones.

Supongamos a continuación que $y > z > x$, una externalidad negativa débil. Los agentes están mejor como resultado de su esfuerzo, pero, debido a los costes que se imponen unos a otros, no tanto como esperaban. La gente que viaja a diario a su trabajo en auto quizás esté mejor de lo que lo estaría si utilizara el transporte público, en caso de que éste se encuentre en malas condiciones, pero los atascos y la contaminación le impiden beneficiarse en la medida esperada. Si la externalidad es un producto de los atascos, es muy difícil que quienes la padecen omitan advertirla. En cambio, si es un producto de la contaminación, podría pasar algún tiempo antes de que las personas entendieran que se perjudican unas a otras, en vez de ser víctimas de (pongamos por caso) la contaminación fabril.

Supongamos para terminar que $y > x > z$, una externalidad negativa fuerte. Todos los agentes están peor como resultado de que todo el mundo trata de estar mejor. Ésta fue una de las principales acusaciones de Marx contra la economía capitalista descentralizada. Su explicación fundamental de las crisis capitalistas, la «teoría de la tendencia decreciente de la tasa de ganancia», tenía esa estructura general. Para mantener o incrementar las ganancias, sostenía Marx, cada capitalista tiene el incentivo del reemplazo de la mano de obra por maquinarias. Sin embargo, cuando todos los capitalistas llevan a cabo ese reemplazo de manera simultánea, serruchan colectivamente la rama en la que están sentados, dado que el origen último de la ganancia es la plusvalía generada por la mano de obra. El argumento

es seductor, pero un análisis más detenido demuestra que es erróneo desde todo punto de vista. Más interesante es otra observación que Marx hizo de pasada y que luego se convirtió en una piedra angular de la teoría del desempleo elaborada por John Maynard Keynes. Cada capitalista, señalaba Marx, tiene una relación ambigua con los trabajadores. Por un lado, quiere que los trabajadores que *él* emplea tengan bajos salarios, pues esto contribuye a elevar las ganancias. Por otro, quiere que todos los *demás* trabajadores tengan salarios altos, dado que esto contribuye a elevar la demanda de sus propios productos. Aunque cualquier capitalista en particular tenga la posibilidad de satisfacer ambos deseos, es lógicamente imposible que todos los capitalistas puedan hacerlo al mismo tiempo. Ésta es una «contradicción del capitalismo» que Keynes expresó del siguiente modo. En una situación de caída de la ganancia, cada capitalista responde con el despido de obreros, y de esa manera ahorra en salarios. No obstante, como la demanda de los trabajadores es lo que directa o indirectamente sostiene las empresas, el efecto del despido simultáneo de obreros por parte de todos los capitalistas será una reducción adicional de las ganancias, que conducirá a más despidos o quiebras.

Hay muchos casos de este tipo general. La sobrepesca, la deforestación y el pastoreo excesivo («la tragedia de los pastos comunes«) pueden ser individualmente racionales, pero menos que óptimos y hasta desastrosos en el plano colectivo. Si cada familia de un país en vías de desarrollo tiene muchos hijos como seguro contra la pobreza en la vejez, la superpoblación generará más pobres. En una crisis de escasez de agua, cada individuo que la utiliza para finalidades no esenciales provoca un leve ascenso en la probabilidad de que las autoridades decidan cortar el suministro durante unas horas por día, con lo cual se afectarán también finalidades esenciales. Estas consecuencias pueden o no ser previstas. Un rasgo crucial de esta categoría de consecuencias indeliberadas es que, aun cuando se las prevea, el comportamiento será el mismo. Como explico en el próximo capítulo, se trata de una *estrategia dominante*: es racional escogerla con prescindencia de lo que hagan los demás.

Internalidades

Un argumento parcialmente similar es válido para las «internalidades», definidas como el beneficio o el perjuicio que la elección de una persona en

un momento dado puede provocar sobre el bienestar que ella misma obtiene en virtud de elecciones posteriores. Metafóricamente hablando, las internalidades son externalidades que una persona impone a sus «*selves* ulteriores». En el examen de la custodia de los hijos resumido en la Figura 11.3, sostuve que el tiempo pasado con el niño genera una internalidad positiva en el progenitor. La adicción representa un ejemplo importante de internalidad negativa. Cuanto mayor haya sido en el pasado su consumo de una sustancia adictiva, menos placentero será para la persona el consumo presente. Este efecto de «tolerancia» también puede producirse con bienes no adictivos. Aunque nos encante el helado de pacana, es probable que nos hartemos si lo tomamos cinco veces por día. En la adicción, sin embargo, el consumo pasado tiene un efecto adicional. Si bien hace que el consumo actual sea menos placentero de lo que sería si el agente no hubiera consumido en el pasado, también aumenta la diferencia de bienestar entre el consumo y el no consumo presentes («retirada»). En la Figura 18.2 se verá una representación esquemática.

Así, independientemente de que se haya abstenido o haya consumido en el pasado, en el presente el agente está mejor si consume que si no consume. El consumo es una estrategia dominante. Al mismo tiempo, un consumo reiterado lo pone en todo momento (salvo unas pocas veces al comienzo) en peor situación que la abstención reiterada, del mismo modo que la estrategia dominante de tener muchos hijos puede empeorar la situación de todos. Hay, desde luego, diferencias evidentes entre las externalidades y las internalidades. Una es la asimetría temporal: en tanto que todos los individuos pueden perjudicarse unos a otros, los *selves* ulteriores no pueden perjudicar a los *selves* anteriores. Otra diferencia es el hecho de que los *selves* sucesivos sólo son en realidad fragmentos temporales de *un único* decisor, mientras que los diferentes individuos no son partes espacialmente distintas de un superorganismo. Una vez que *la* persona (la única) comprende que sus elecciones presentes ejercen un efecto negativo sobre el bienestar que puede obtener en virtud de elecciones posteriores, tiene un incentivo para modificar su comportamiento. Que el incentivo tenga la fuerza suficiente dependerá de la gravedad de los síntomas de retirada y de la magnitud del descuento que el agente haga del bienestar futuro. Algunos agentes que nunca habrían dado el primer paso de haber conocido las consecuencias tal vez decidan no abandonar una vez que están enganchados.

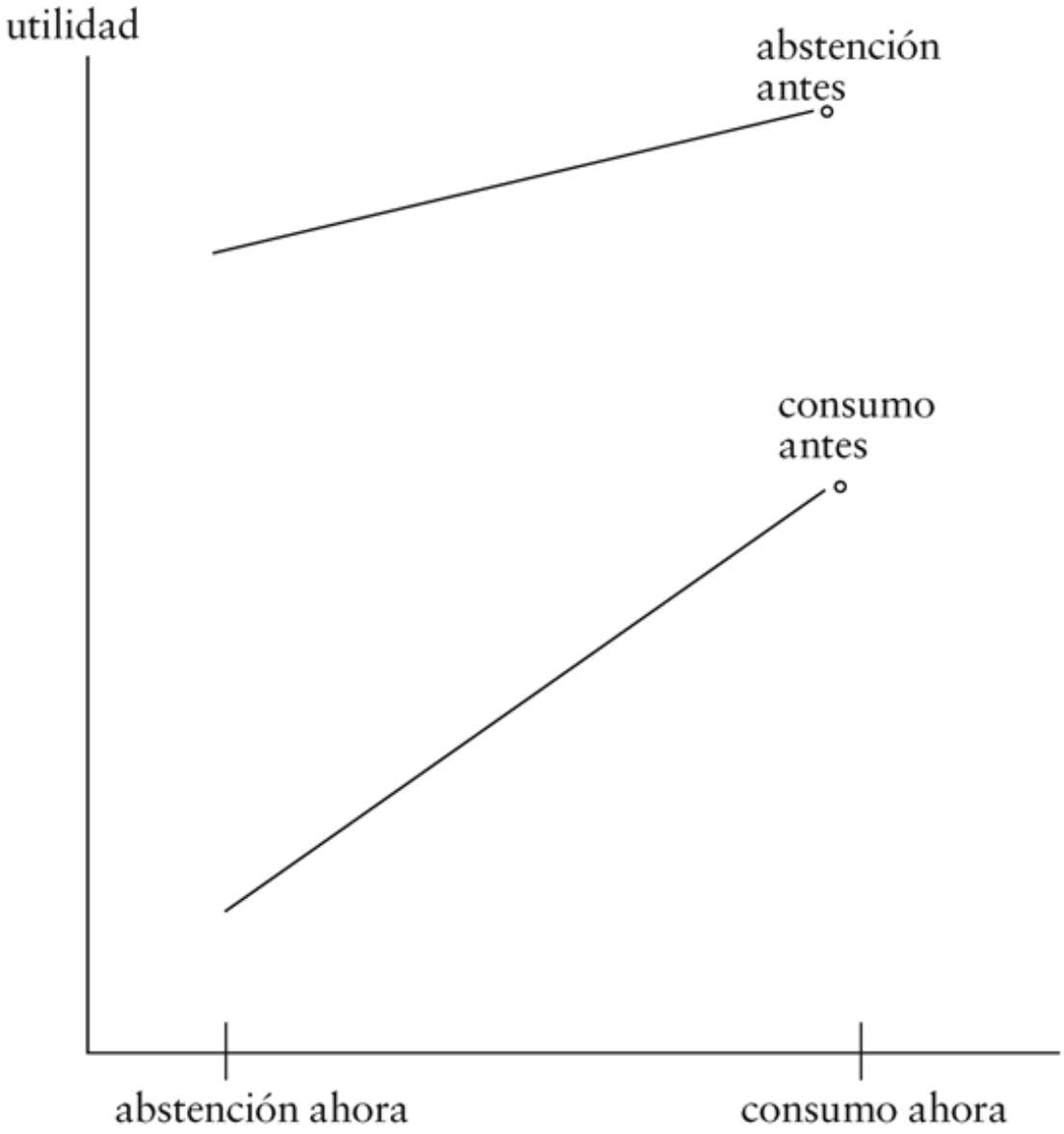


FIGURA 18.2

El síndrome del hermano menor

Las consecuencias indeliberadas de la acción social también pueden ser producto de lo que llamaré *mecanismo del hermano menor*. Antes de explicar esta expresión, querría ilustrarla con un famoso ejemplo tomado de la teoría económica, la «telaraña», también denominada «ciclo del cerdo» porque se propuso por primera vez como explicación de las fluctuaciones

cíclicas de la producción porcina. Sin embargo, tiene una aplicación mucho más amplia. Las fluctuaciones de la industria naval han mostrado con frecuencia el mismo patrón, con un mercado favorable a los vendedores seguido por una sobreinversión y un exceso de oferta. Cuando los estudiantes eligen su carrera sobre la base de la demanda actual de graduados, pueden desmoronar colectivamente el fundamento de sus decisiones.

Los criadores de cerdos deben decidir con un año de anticipación la cantidad de ejemplares que quieren poner en el mercado el año siguiente, una decisión determinada por el precio al que esperan vender los animales y el coste de producirlos. Un incremento en el precio esperado impulsará a los criadores a producir más, según lo refleja la curva ascendente de la oferta en la Figura 18.3.

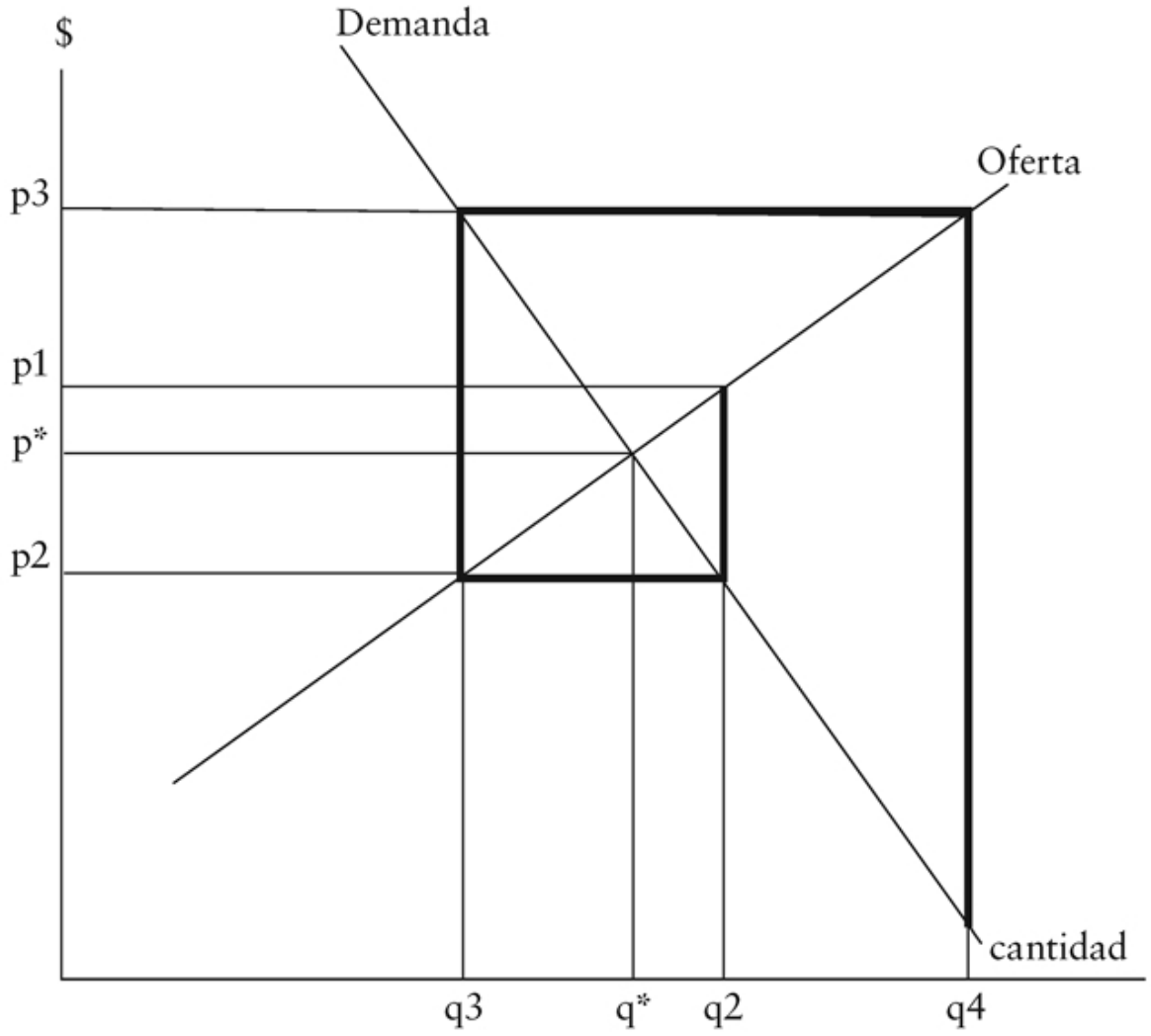


FIGURA 18.3

Supongamos que en el año 1 el precio de los cerdos es p1. Como esperan que los precios se mantengan en el año 2, ese año los criadores destinan al mercado una cantidad q2. En esos volúmenes, sin embargo, el precio de equilibrio del mercado es p2 y no p1. Con la expectativa de que los precios se mantendrán en ese nivel en el año 3, los criadores producen un volumen q3 para ese período. El precio de equilibrio del mercado será p3, lo cual los inducirá a producir q4 en el año 4, y así sucesivamente. En este caso, los precios y los volúmenes forman un patrón de espiral hacia adentro o «telaraña» indicado en el diagrama por las líneas en negrita. Las sorpresas agradables alternan con las desagradables, pero el resultado

esperado nunca aparece. Si los gradientes relativos de las curvas de oferta y demanda se modifican, el resultado podría ser una espiral hacia afuera convergente en el precio de equilibrio p^* y el volumen de equilibrio q^*.

Hay algo irracional en la conducta de los criadores. *Cada uno* de ellos cree tener la libertad de variar su producción para maximizar sus ganancias, y a la vez se forma el supuesto tácito de que los otros se limitan a hacer mecánicamente lo que han hecho el año anterior. Si bien quizás irracional, el comportamiento es sin duda inteligible. Un filósofo francés, Maurice Merleau-Ponty, ha dicho que nuestra tendencia espontánea consiste en ver a otras personas como «hermanos menores» [*«younger siblings»*].[6] No nos resulta fácil atribuir a otros la misma capacidad de deliberación y reflexión que, según nos dice la introspección, nosotros mismos poseemos, y tampoco, ya que estamos, nuestra confusión interna, nuestras dudas y nuestras angustias (véase el Capítulo 23). La idea de ver en los otros a personas tan estratégicas y calculadoras como nosotros no parece surgir naturalmente.

Tres ejemplos del comportamiento electoral también pueden ilustrar la idea. Supongamos que soy miembro del ala izquierda del Partido Socialista de mi país. Preferiría con mucho que en el poder estuvieran los socialistas y no los comunistas, pero como las encuestas pronostican una sólida mayoría socialista, voto a los comunistas para impulsar a mi partido a desplazarse hacia la izquierda. Sin embargo, no me detengo a preguntarme si otros simpatizantes del ala izquierda hacen el mismo razonamiento. Si muchos de ellos lo hacen, los comunistas podrían ganar. La intención de producir el resultado de primer orden (una victoria socialista con una buena actuación comunista) puede generar el resultado de tercer orden (una victoria comunista). En lo que es probablemente un escenario más habitual, si muchos votantes se quedan en sus casas porque confían en que su partido ganará, éste tal vez pierda. Para terminar, recuérdese el ejemplo de Chirac del Capítulo 12. Una explicación posible de su desastrosa convocatoria a elecciones anticipadas puede ser el hecho de que omitiera prever que los votantes se basarían en esa decisión para inferir sus creencias, en vez de comportarse simple y mecánicamente como decían que lo harían en las encuestas.

La incapacidad de ver a otros como agentes intencionales y maximizadores se observa cuando los legisladores o administradores proponen políticas que pierden parte de su fuerza cuando los agentes se

ajustan a ellas. Según el derecho romano, el robo de un solo caballo o buey convertía al hombre en ladrón de ganado, mientras que no era delito el robo de menos de cuatro cerdos o diez ovejas. Un comentarista señaló que «en tal situación jurídica, cabría esperar que los robos de tres cerdos u ocho ovejas fueran anormalmente comunes». Para promover la estabilidad laboral, muchos países han sancionado leyes que disponen la ilegalidad del despido de trabajadores con, digamos, dos años o más de antigüedad en su puesto. La respuesta racional de los empleadores consiste en tercerizar u ofrecer contratos de trabajo temporarios, con lo cual se reduce la estabilidad laboral. Las ciudades pueden construir autopistas para reducir los atascos, con el único resultado de comprobar que, al aumentar la cantidad de gente que se traslada en automóvil a su trabajo, en las calles hay tantos embotellamientos como antes y la contaminación es más grande. El Gobierno puede tratar de limitar la inmigración a quienes están casados con una persona que ya tiene residencia legal en el país, pero el efecto será inducir a la gente a casarse con ese único objeto. Las exenciones al servicio militar que benefician a los estudiantes crean un incentivo para ir a la universidad. Aun la creación de sistemas explícitos de incentivos puede resultar contraproducente. Así, cuando la Administración de Escuelas Públicas de Chicago estableció un sistema que aplicaba diversas sanciones a las escuelas de mal desempeño, a sus creadores no se les ocurrió que, además de introducir un incentivo para que los docentes se esforzaran más en su trabajo, el sistema los alentaba a inflar las notas, como efectivamente sucedió.

El síndrome del hermano menor puede tener importantes consecuencias sociales, como lo revelarán dos ejemplos. Tocqueville señala que en las décadas previas a la Revolución francesa, las clases altas denunciaron públicamente los vicios del régimen y su devastador impacto sobre la población, como si ésta fuera sorda a lo que decían. «Esto me recuerda la impresión de Madame Duchâtelet, a quien, según el secretario de Voltaire, no le importaba desvestirse delante de sus criados, pues no estaba convencida de que los *valets* fueran hombres.» En sus memorias, el secretario escribió en realidad que las «grandes damas sólo miraban a sus lacayos como autómatas». Esta muestra simultánea de desdén hacia las clases bajas y de denuncia de su infortunio preparó las mentes para la revolución. Más recientemente, el argumento subyacente a la «curva de Phillips», conforme con el cual el Gobierno puede decidir, si así lo desea,

alcanzar un bajo índice de desempleo a costa de una elevada inflación, presupone que los actores sociales son ignorantes de esta política. Sin embargo, cuando los gobiernos procuraron alcanzar esa meta, el comportamiento estratégico de sindicatos y otros actores racionales socavó sus esfuerzos y produjo, en cambio, «estanflación»: elevada inflación *y* elevado desempleo.

A diferencia de las consecuencias indeliberadas generadas por las externalidades, las producidas por el mecanismo del hermano menor pueden terminar cuando los agentes lo entienden. En los casos que he descrito no hay estrategias dominantes, sólo estrategias que son óptimas bajo el supuesto (habitualmente implícito) de que los otros son menos racionales que uno. Una vez que todos los agentes se ven de manera recíproca como racionales, su comportamiento puede converger en un resultado plenamente predecible. Todos los criadores de cerdos esperarán que prevalezca el precio de equilibrio. Al actuar sobre la base de esa expectativa, producirán el volumen de equilibrio. Su creencia compartida se cumple por sí misma. Esta idea es el tópico del siguiente capítulo.

* * *

NOTA BIBLIOGRÁFICA

El impacto de la adicción en el descuento temporal está documentado en Louis A. Giordano y otros, «Mild opioid deprivation increases the degree that opioid-dependent outpatients discount delayed heroin and money», *Psychopharmacology*, 163(2), 2002, págs. 174-182. La nota que propone un modelo del cuento de Andersen está inspirada en Carl Christian von Weiszäcker, «Notes on endogenous change of tastes», *Journal of Economic Theory*, 3(4), diciembre de 1971, págs. 345-372. Se encontrará un análisis de las ideas de Marx sobre las consecuencias indeliberadas en mi *Making Sense of Marx*, Cambridge y Nueva York, Cambridge University Press, 1985, Capítulo 1.3.2 y pássim. El modelo de la adicción procede de Gary Becker y Kevin Murphy, «A theory of rational addiction», *Journal of Political Economy*, 96(4), agosto de 1988, págs. 675-700. Thomas C. Schelling, *Micromotives and Macrobehavior*, Nueva York, Norton, 1978 [*Micromotivos y macroconducta*, México, Fondo de Cultura Económica, 1989], es un espléndido estudio conceptual de las consecuencias indeliberadas. Con respecto a la idea de las internalidades, véase Richard J.

Herrnstein y otros, «Utility maximization and melioration: internalities in individual choice», *Journal of Behavioral Decision Making*, 6(3), 1993, págs. 149-185. El ejemplo de la Administración de Escuelas Públicas de Chicago se ha tomado de Brian A. Jacob y Steven D. Levitt, «Rotten apples: an investigation of the prevalence and predictors of teacher cheating», *The Quarterly Journal of Economics*, 118(3), 2003, págs. 843-877.

Notas al pie

[1] En términos más formales, imaginemos una persona que de manera regular (pero no consciente) ajusta sus deseos a fin de tener mayor preferencia por la mercancía que en ese momento más escasea en sus existencias. Supongamos que está expuesta a la siguiente secuencia de grupos de dos mercancías: (1/2, 3/2), (3/4, 1/2), (1/4, 3/4), (3/8, 1/4), … Luego, si en un momento dado consume el grupo *n* de la secuencia y para el período siguiente se le ofrece la elección entre el grupo *n* y el grupo *n* + 1, siempre elegirá este último, porque ofrece más de la mercancía de la que actualmente la persona menos tiene. Pero como la secuencia converge hacia cero, estas mejoras locales le acarrean la ruina global. El efecto es similar al de las preferencias cíclicas que pueden provocar una sangría mortal en un agente (Capítulo 11), pero el mecanismo es diferente.

[2] De manera accidental, la expresión «efecto de dotación» ha llegado a utilizarse tanto para la tendencia a sobrevaluar artículos en nuestra posesión como para la utilidad que una persona puede extraer en el presente de la utilidad en el pasado (Capítulo 2). No hay relación alguna entre los dos significados.

[3] A decir verdad, cuesta entender cómo podría prevérselo. Para ello habría que utilizar simultáneamente dos puntos iniciales, el estado previo a la compra en relación con el cual el bien es una ganancia y el estado posterior a la compra en relación con el cual es una pérdida.

[4] Esta perspectiva limitada es compartida por algunos especialistas en ciencias sociales, quienes sostienen que emociones como la ira pueden ser «racionales», o al menos adaptativas, porque permiten a los agentes salirse con la suya en encuentros con otros.

[5] Muchos economistas no tendrían en cuenta como externalidades todos los fenómenos que enumero aquí. Incluirían la contaminación, pero no efectos generados por el mercado como el desempleo keynesiano. En lo que me concierne, sin embargo, lo importante es lo que tienen en común: en la búsqueda de un beneficio para sí, cada individuo impone un pequeño coste o beneficio a todos los demás y *a sí mismo*. Una empresa que despida trabajadores o recorte salarios causará una pequeña reducción de la demanda de sus propios productos.

[6] En realidad escribió «hermano menor» [*«younger brother»*].*

* *Siblings*, como escribe el autor, es el sustantivo para designar a los hermanos sin distinción de género. [T.]

Capítulo 19

Interacción estratégica

Interacción estratégica con elecciones simultáneas

La invención de la *teoría de los juegos* puede llegar a verse como el avance más importante de las ciencias sociales en el siglo XX. Su valor es en parte explicativo, pero principalmente conceptual. En algunos casos, la teoría nos permite explicar comportamientos que con anterioridad parecían desconcertantes. Pero lo más importante es que esclarece la estructura de la interacción social. Una vez que vemos el mundo desde el prisma de la teoría de los juegos –o la «teoría de las decisiones interdependientes», como quizá sería más adecuado llamarla–, nada vuelve a parecer del todo igual.

Considero en primer lugar los juegos en los que los agentes toman decisiones simultáneas. La meta es entender si y cómo n agentes o *jugadores* pueden alcanzar una coordinación no impuesta de sus *estrategias*. Con frecuencia, tomaremos en cuenta el caso especial de $n = 2$. Los jugadores tal vez sean capaces de comunicarse entre sí, pero no de concertar acuerdos vinculantes. Para cualquier *n-tupla* de estrategias, una escogida por cada agente, corresponde un *resultado*. Cada agente clasifica los posibles resultados de acuerdo con su *orden de preferencias*. Cuando sea necesario, supondremos que se cumplen las condiciones para representar las preferencias como utilidades cardinales (Capítulo 11). La *estructura de recompensas* es la función que a cualquier *n-tupla* de estrategias asigna una *n-tupla* de utilidades. Aunque la palabra «recompensa» pueda sugerir un resultado monetario, la palabra se utilizará para referirse a resultados

psicológicos (utilidades y, en última instancia, preferencias). Cuando, como sucede a menudo, hay una divergencia entre la estructura de recompensas monetarias o materiales y la estructura de recompensas psicológicas, sólo esta última es relevante.

Como se mencionó brevemente en el capítulo anterior, un agente puede tener una estrategia que es *dominante* en el sentido de que, hagan los otros lo que hicieren, genera para él un resultado más favorable que el que obtendría si escogiera cualquier otra estrategia. Su *resultado* acaso dependa de lo que otros hacen, pero su *elección* no. En otros casos hay una auténtica interdependencia de elecciones. Si los otros conducen por el lado izquierdo del camino, mi mejor respuesta es conducir también por ese lado; si conducen por la derecha, mi mejor respuesta es conducir por la derecha.

Un *equilibrio* es una *n-tupla* de estrategias con la propiedad de que ningún jugador puede, si se aparta de su estrategia de equilibrio, producir de manera unilateral un resultado que prefiera estrictamente al resultado de equilibrio. De modo equivalente, en el equilibrio la estrategia elegida por cada jugador es la mejor respuesta a las estrategias elegidas por los otros, en el sentido débil de que él no puede hacer *nada mejor* que escoger su estrategia de equilibrio si los otros escogen la suya. Sin embargo, no es necesario que la estrategia sea óptima en el sentido fuerte de que el jugador estará *peor* si se desvía unilateralmente. En el caso general, un juego puede tener varios equilibrios. Pronto veremos algunos ejemplos. Supongamos, con todo, que hay un solo equilibrio. Supongamos además que la estructura de recompensas y la racionalidad de todos los jugadores son de conocimiento general.[7] Con estos supuestos, podemos predecir que todos los agentes escogerán su estrategia de equilibrio, dado que es la única basada en creencias racionales sobre lo que harán los otros.

Algunos juegos con un único equilibrio giran en torno de la existencia de estrategias dominantes. La frase «giran en torno de la existencia de estrategias dominantes» puede significar una de dos cosas, ilustradas en los paneles A y B de la Figura 19.1.[8] En un accidente en el que intervienen dos automóviles, ambos sufren daños. En un accidente en que intervienen un peatón y un auto, sólo el primero sufre daños. Los accidentes entre dos autos ocurren si al menos uno de los conductores es imprudente. Si ambos lo son, el resultado es peor. Los accidentes entre un automóvil y un peatón ocurren sólo si ambos son imprudentes. Ser prudente como es debido es costoso. De estas premisas se deduce que, en el caso de los dos autos, ser

prudente es la estrategia dominante para cada uno de los conductores. En el caso del auto y el peatón, la falta de prudencia es dominante para el conductor. El peatón no tiene estrategia dominante, dado que la debida prudencia es la mejor respuesta a la imprudencia y la imprudencia es la mejor respuesta a la debida prudencia. Como sabe que el conductor tiene como estrategia dominante la imprudencia y, por ser racional, la escogerá, el peatón elegirá, no obstante, la debida prudencia.[9]

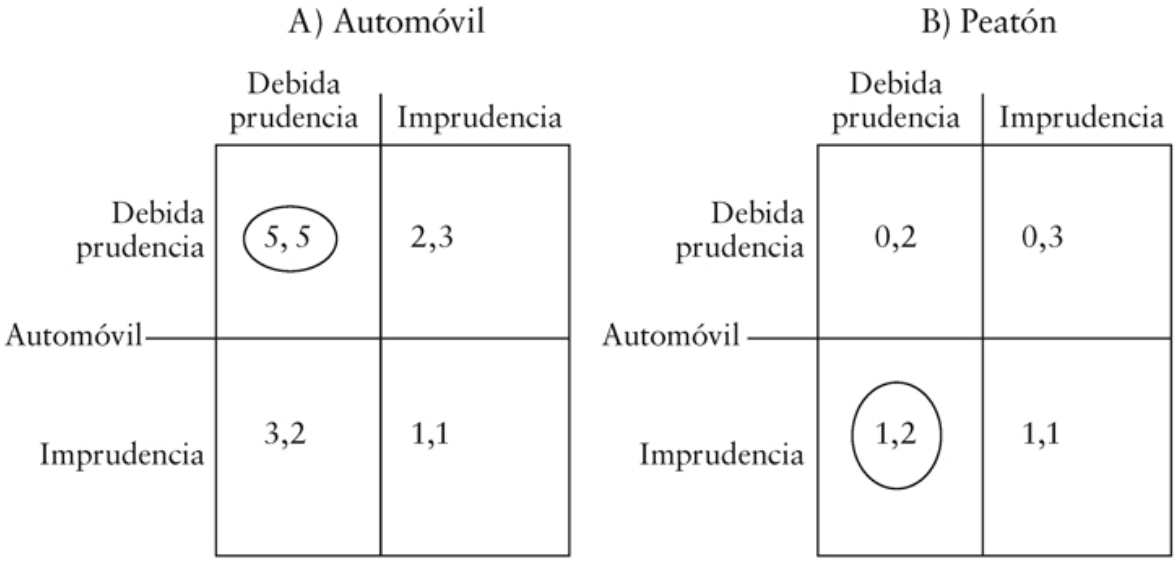


FIGURA 19.1

Los juegos en los que todos los jugadores tienen estrategias dominantes son bastante comunes y, como veremos, empíricamente importantes. Desde un punto de vista teórico son un tanto triviales, salvo cuando se repiten a lo largo del tiempo. Los juegos en los que algunos jugadores tienen estrategias dominantes que pueden inducir elecciones bien definidas en otros son menos comunes, pero también importantes. Sin embargo, plantean exigencias informacionales más grandes, pues en nuestro ejemplo el peatón necesita conocer los posibles resultados para el conductor, al igual que para sí mismo, mientras que los dos conductores sólo necesitan conocer sus propios resultados. Con frecuencia, podemos atribuir estrategias dominantes a otros sin muchas dificultades. Por ejemplo, no solemos mirar hacia ambos lados antes de cruzar una calle de sentido único, porque suponemos que el

temor de los conductores a ser responsables de un accidente los llevará a respetar la regla que fija ese sentido único.

Una clase especial de juegos tiene *equilibrios de coordinación*, a menudo denominados «convenciones», en los que cada jugador no sólo no tiene incentivos para desviarse en forma unilateral, sino que también preferiría que ningún otro lo hiciera. En un equilibrio en el cual todo el mundo conduce por el lado derecho del camino, podría ocurrir un accidente si yo me desviara *o* si alguna otra persona lo hiciera. En este caso, el equilibrio no es único, dado que la conducción por el lado izquierdo tiene las mismas propiedades.[10] Muchas veces, no importa lo que hagamos con tal de que todos hagamos lo mismo. Los significados de las palabras son arbitrarios, pero una vez que se fijan, se transforman en convenciones. En otros casos, sí importa lo que hagamos, pero es más importante que todos hagamos lo mismo. A la brevedad volveré a algunos ejemplos.

Dos ejemplos de duopolio

Algunos juegos tienen equilibrios únicos que no giran alrededor de la existencia de estrategias dominantes. El comportamiento de los duopolios es un ejemplo (véase la Figura 19.2). Cuando dos empresas dominan un mercado, la baja de la producción de una de ellas provocará la elevación de los precios y una expansión de la producción de la otra. En otras palabras, cada empresa tiene un esquema de «mejor respuesta» que le dice cuánto producir en función de lo producido por la otra. En equilibrio, la producción de cada una de ellas es la mejor respuesta a la producción de la otra. Este enunciado no implica que no puedan hacer algo mejor. Si constituyeran un cártel y restringieran su producción a niveles por debajo del equilibrio, ambas obtendrían grandes ganancias. No obstante, estos niveles de producción colectivamente óptimos no son las mejores respuestas recíprocas. De hecho, las empresas enfrentan un dilema del prisionero (definido en la Figura 19.3).

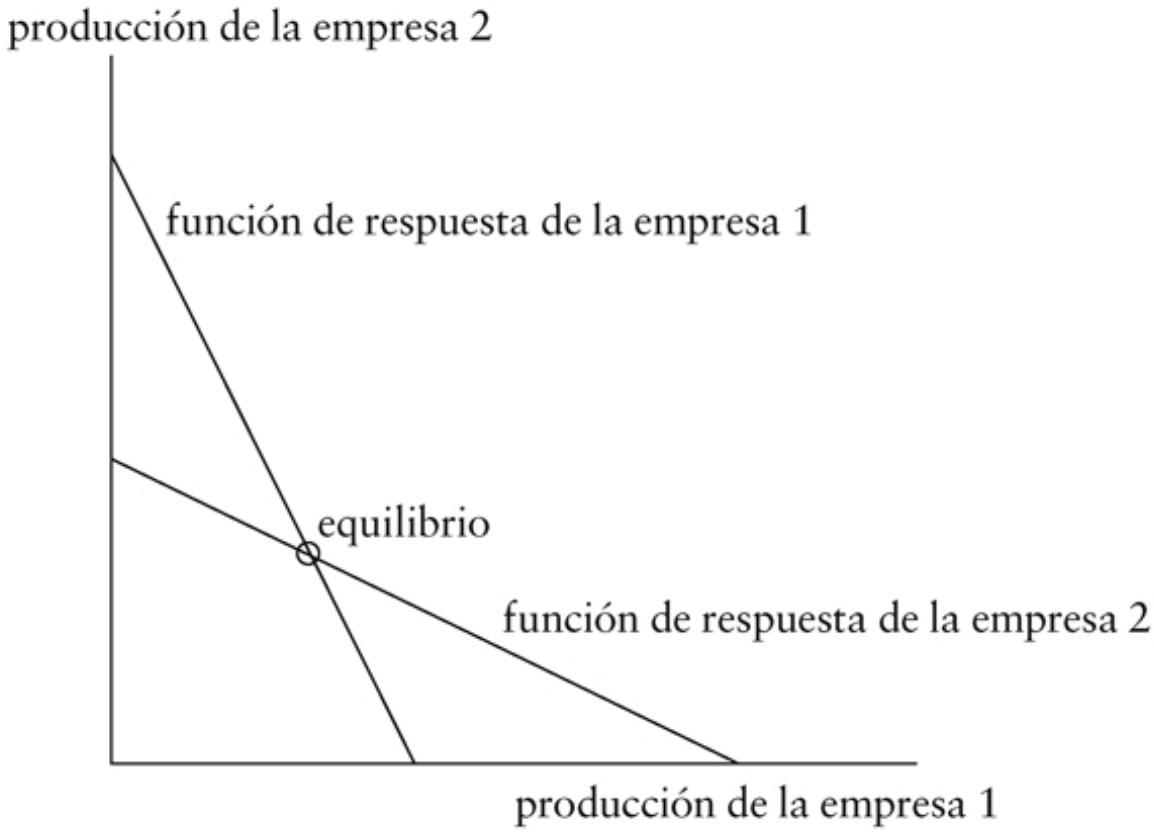


FIGURA 19.2

Para ver otro caso de duopolio, consideremos a dos vendedores de helados en una playa, que tratan de encontrar la mejor ubicación para sus quioscos, con el supuesto de que los clientes (de quienes se presume que están distribuidos de manera pareja a lo largo de la costa) irán al más cercano. No hay una estrategia dominante. Si uno de ellos pone un quiosco a cierta distancia del centro de la playa, hacia la izquierda, la mejor respuesta del otro consistirá en ubicarse inmediatamente a la derecha, ante lo cual la mejor respuesta del primero será moverse de nuevo a la derecha, y así sucesivamente, hasta que sus quioscos queden uno al lado del otro en el medio de la playa. Como es obvio, este equilibrio único no es el mejor para el conjunto de los clientes. Para ellos, el mejor resultado es aquel en el cual cada quiosco esté ubicado a mitad de camino entre el centro y un extremo de la playa. Aunque este resultado es tan conveniente para los vendedores como el resultado de equilibrio, esas ubicaciones no son las mejores respuestas recíprocas. Este modelo también se ha aplicado para explicar la

tendencia de los partidos políticos (en un sistema bipartidista) a moverse hacia el centro del espectro político.

Supongamos, sin embargo, que cuando ambos quioscos están en el medio, los clientes cercanos a los extremos se abstienen de comprar helados porque, de hacerlo, éstos se derretirán en el camino de vuelta. Si ningún cliente está dispuesto a caminar más de la mitad de la longitud de la playa, un cuarto para llegar al quiosco y un cuarto para volver, el resultado óptimo para el consumidor también es el equilibrio único, pues ninguno tiene incentivos para cambiar de ubicación. Supongamos que la playa tiene mil metros de largo. Si el vendedor que está a setecientos cincuenta metros muda su quiosco hasta los setecientos, perderá los cincuenta clientes situados entre los novecientos cincuenta y los mil que no están dispuestos a caminar más de cuatrocientos cincuenta, y ganará los veinticinco clientes situados entre los cuatrocientos setenta y cinco y los quinientos, para quienes su quiosco está ahora más cerca que el otro: una pérdida neta. Un argumento similar podría explicar también por qué los partidos nunca convergen plenamente en el centro, dado que los extremistas de uno y otro lado preferirían abstenerse antes que votar por un partido centrista. Además, como he señalado al final del Capítulo 17, sencillamente no es verosímil considerar que la maximización del voto es el único objetivo de los partidos políticos.

Algunos juegos frecuentes

Unas cuantas estructuras de interacción simples, con retribuciones como las mostradas en la Figura 19.3, aparecen con mucha frecuencia en una gran variedad de contextos.[11] C y D significan «cooperación» y «defección». En el juego del teléfono, el jugador de las columnas ha sido el primero en llamar. En el juego del punto focal, A y B pueden ser cualquier par de acciones tales que ambos jugadores prefieran coordinarse en una u otra antes que no coordinarse, pero las dos maneras de hacerlo les dan igual.

Los juegos esclarecen la estructura de los dos problemas centrales de la interacción social, la *cooperación* y la *coordinación*. En una sociedad sin cooperación para beneficio mutuo, la vida sería «solitaria, pobre, detestable, brutal y breve» (Hobbes). El hecho de que sería *previsiblemente* mala es un magro consuelo. En una sociedad donde la gente fuera incapaz de coordinar su conducta, las consecuencias indeliberadas abundarían y la vida sería

como «un cuento contado por un idiota, lleno de ruido y de furia y que no significa nada» (*Macbeth*). La cooperación y la coordinación a veces son fructíferas, pero a menudo fracasan estrepitosamente. La teoría de los juegos puede dilucidar tanto los éxitos como los fracasos.

	C	D
C	3,3	0,4
D	4,0	(1.5, 1.5)

Dilema del prisionero

	C	D
C	(4,4)	1,3
D	3,1	(2,2)

Cacería del ciervo/
juego de la seguridad

	C	D
C	2,2	(1,3)
D	(3,1)	0,0

Gallina

	Ballet	Box
Ballet	(1,2)	0,0
Box	0,0	(2,1)

Batalla de los sexos

	Volver a llamar	No volver a llamar
No volver a llamar	(2,2)	0,0
Volver a llamar	0,0	1,1

Juego del teléfono

	A	B
A	1,1	0,0
B	0,0	(1,1)

Juego de punto focal

FIGURA 19.3

El dilema del prisionero (DP), la cacería del ciervo y la gallina implica de una manera u otra la elección entre la cooperación y la defección (no cooperación). El dilema del prisionero tiene ese nombre porque en un primer momento se utilizó la siguiente historia para ilustrarlo. A cada uno de dos prisioneros, que han participado en el mismo delito pero están ahora en celdas separadas, se le dice que si delata al otro y éste no lo delata a él, saldrá libre y el otro irá a prisión durante diez años; si no se delatan mutuamente, ambos pasarán un año en la cárcel, y si se delatan uno a otro, recibirán una sentencia de cinco años de prisión.[12] En esas circunstancias, la delación es una estrategia dominante, aunque ambos estarían mejor si ninguno delatara. El resultado es el producto de una combinación de la «tentación del aprovechado» (salir en libertad) y el «miedo a ser embaucado» (recibir una sentencia de diez años).

Las externalidades negativas examinadas en el último capítulo también pueden verse como DP de muchas personas. Siguen algunos otros ejemplos. Para cada trabajador (si suponemos motivaciones egoístas), es mejor no estar sindicalizado que afiliarse a un sindicato, aun cuando para la totalidad sea mejor que todos estén afiliados y obtengan salarios más altos. Para cada empresa integrante de un cártel, es mejor desvincularse y producir un volumen elevado para aprovechar los altos precios generados por las restricciones a la producción en las demás firmas, pero cuando todas hacen lo mismo, los precios caen al nivel competitivo; la maximización de las ganancias de cada empresa socava la maximización de las ganancias conjuntas. El cártel de la Organización de Países Exportadores de Petróleo (OPEP) es vulnerable justamente en ese aspecto. Otros ejemplos son las situaciones en las que todo el mundo tiene que correr lo más rápidamente posible para permanecer en el mismo lugar, como la carrera armamentística entre los Estados Unidos y la ex Unión Soviética, la publicidad política o los estudiantes que escriben ensayos para un docente que «se basa en el promedio para calificar».

La idea de la cacería del ciervo se suele atribuir a Jean-Jacques Rousseau, aunque su lenguaje era un tanto opaco.[13] En una forma más estilizada, este juego implica la participación de dos cazadores que pueden elegir entre cazar un ciervo (C) o una liebre (D). Cada uno de ellos puede

cazar una liebre por sí solo, pero es necesario (y suficiente) el esfuerzo conjunto de ambos para cazar un ciervo. Medio ciervo vale más que una liebre. La cacería de liebres exige más tiempo y esfuerzo cuando los dos tratan de cazarlas, porque sus ruidos las ahuyentan. Como en el caso del dilema del prisionero, se corre el riesgo de pasar por memo y procurar cazar un ciervo, cuando el otro va en busca de una liebre. Sin embargo, no existe la tentación del aprovechado. El juego tiene dos equilibrios, en los casilleros superior izquierdo e inferior derecho.

Aunque el primer equilibrio es notoriamente mejor, tal vez no se concrete. Para ver por qué podría suceder así, podemos dejar de lado el supuesto de que la estructura de retribuciones es de conocimiento general y permitir que los agentes tengan creencias erróneas sobre esa estructura en otros agentes. Las acciones emprendidas sobre la base de esas creencias constituirán un *equilibrio en sentido débil* si, en el caso de cada agente, las acciones de los otros confirman sus creencias sobre ellos. Supongamos, por ejemplo, que en una cacería del ciervo cada agente cree falsamente que los demás tienen preferencias del DP. Dada la creencia, la acción racional es desertar, y confirmar así la creencia de los otros de que *él* tiene esas preferencias. Esta sociedad podría terminar con elevados niveles de evasión impositiva y corrupción. Vuelvo a esos casos de «ignorancia pluralista» en el Capítulo 23. En otra sociedad, donde la gente acierta al creer que los otros tienen preferencias de la cacería del ciervo, surgirá un buen equilibrio en el cual todos pagan sus impuestos y no ofrecen ni aceptan sobornos. Las «culturas de la corrupción» podrían ser un fenómeno dependiente de las creencias y no de las motivaciones.

El control internacional de enfermedades infecciosas puede tener la estructura de una cacería del ciervo. Si un solo país omite tomar las medidas apropiadas, otros no podrán protegerse.[14] Otro ejemplo es el de las medidas contra el terrorismo. Si sólo una de dos naciones invierte en ellas, beneficia a la otra además de beneficiarse a sí misma. Si los costes exceden los beneficios que obtiene, no invertirá en forma unilateral. Sin embargo, si ambas invierten, la capacidad de compartir la información puede llevar a cada una a un nivel de seguridad más grande que el que podría alcanzar si explotara la inversión de la otra.

En estos ejemplos, la estructura de retribuciones surge de la naturaleza causal de la situación. En la cacería del ciervo y el caso del control de enfermedades, el «umbral tecnológico» implica que los esfuerzos

individuales son inútiles. En el caso del contraterrorismo, la causa subyacente es algo semejante a las economías de escala: diez unidades de esfuerzo tienen un efecto más de dos veces más grande que cinco unidades. En otros casos, la estructura de retribuciones se debe al hecho de que los agentes se preocupan por otras cosas al margen de sus propias recompensas materiales. En esas circunstancias, es más habitual hablar de un juego de la seguridad (JS). Aun cuando la estructura de retribuciones materiales es la de un DP, cada individuo puede estar dispuesto a cooperar si está *seguro* de que otros lo harán. El deseo de ser justo, o la renuencia a actuar como un aprovechado, pueden vencer la tentación de explotar la cooperación de los otros. O bien las preferencias altruistas pueden transformar un DP en un JS.

Interpretemos las retribuciones en el DP de la Figura 19.3 como recompensas monetarias y supongamos que la utilidad de cada persona es igual a su recompensa monetaria más la mitad de la recompensa monetaria de la otra. En ese caso, la retribución de la utilidad será como la muestra la Figura 19.4: un JS. El DP también puede transformarse en un JS por medio de un tercer mecanismo, si desde afuera se asocia una penalidad a la elección de la estrategia no cooperativa D. Si interpretamos una vez más las retribuciones en el DP de la Figura 19.3 como recompensas monetarias, *y* suponemos que eso es todo lo que interesa a los agentes, la deducción de 1,25 de la recompensa a la defección lo convertirá en un JS. Un sindicato podría, por ejemplo, imponer sanciones formales o informales a los trabajadores no sindicalizados. Para terminar, podríamos transformar un DP en un JS si recompensáramos la cooperación, con el ofrecimiento, por ejemplo, de una bonificación o un soborno de 1,25 a los cooperadores. Sin embargo, las promesas de recompensa deben respetarse, mientras que una amenaza no tiene que llevarse a la práctica si funciona. Si la retribución de los aprovechados es muy elevada, los beneficios de la cooperación tal vez no sean lo bastante grandes para financiar los sobornos.[15] En algunos casos, no obstante, se utilizan recompensas. Los trabajadores que se afilian a un sindicato quizá se beneficien no sólo con salarios más altos, que por lo común corresponden igualmente a quienes no están sindicalizados, sino también con planes de jubilación y vacaciones baratas ofrecidas sólo a los afiliados.

El juego de la gallina recibe ese nombre de un ritual adolescente mostrado en el filme *Rebelde sin causa*, de 1955. Adolescentes de Los Ángeles llevan automóviles robados hasta un acantilado y participan de un

juego en el cual dos muchachos conducen simultáneamente sus autos hasta el borde de aquél y se detienen en el último momento posible. El primero en detenerse es un «gallina» y pierde. En otra variante, dos autos corren uno hacia otro y el «gallina» es el primero en desviarse. En cada uno de los dos equilibrios, cada agente hace lo opuesto del otro. Aun con un conocimiento general de la estructura de retribuciones y de la racionalidad del agente, no podemos predecir cuál de los equilibrios (si acaso) será el elegido. Desde el punto de vista de la elección racional, la situación es *indeterminada*. En la segunda versión del juego (la del «desvío»), un jugador podría intentar romper la indeterminación si se vendara los ojos (a la vista de todos), e indujera de tal modo al otro a desviarse. No obstante, el escenario así generado sería el mismo, con el solo reemplazo de la opción de «desviarse» o «no desviarse» por la de «vendarse» o «no vendarse».[16] Es una situación profundamente frustrante.

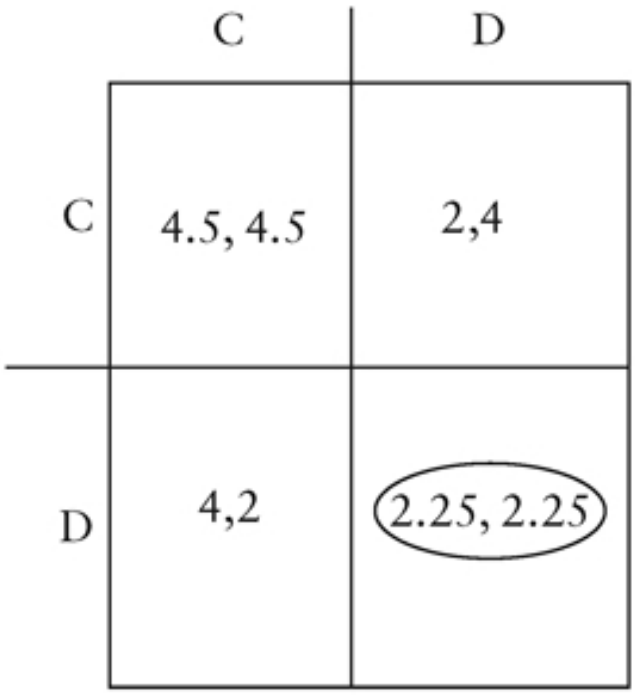


FIGURA 19.4

Según una manera de ver la carrera armamentística, ésta tiene la estructura de la gallina. A menudo se alude a la crisis de los misiles en Cuba como un caso en el cual las dos superpotencias estaban encerradas en

una confrontación al estilo del juego de la gallina, y la URSS fue «la primera en parpadear». Otro ejemplo es el de dos agricultores que utilizan el mismo sistema de riego para sus campos. El sistema puede ser adecuadamente mantenido por una persona, pero los dos granjeros obtienen iguales beneficios de él. Si uno de ellos no cumple con su parte del mantenimiento, el otro quizá tenga, con todo, interés en hacerlo. El caso de Kitty Genovese también puede verse desde esta óptica, de suponerse que cada vecino preferiría intervenir si y sólo si ningún otro lo hiciera.

Para ocuparnos ahora de las cuestiones de *coordinación*, será adecuado considerar en primer lugar la batalla de los sexos. El estereotipo subyacente a la historia es el siguiente. Un hombre y su mujer quieren salir a la noche. Han decidido ir o bien a un ballet o a un encuentro de boxeo luego del trabajo, y han acordado tomar la decisión final por teléfono. Sin embargo, el teléfono del marido está descompuesto, por lo cual tienen que decidir mediante una coordinación tácita. Los une el interés común en estar juntos, pero los separan intereses divergentes acerca del lugar adonde ir. Al igual que el juego de la gallina, este juego tiene dos equilibrios, coordinar la ida al ballet o a la pelea de boxeo. Y como en aquél, no hay manera como el conocimiento general de la estructura de retribuciones y de la racionalidad pueda decir a la pareja dónde encontrarse. Una vez más, la situación es indeterminada.

Juegos de este tipo surgen cuando la coordinación puede adoptar muchas formas, todas las cuales son mejores para los agentes que ninguna coordinación en absoluto, aunque algunos agentes prefieran una de ellas a todas las demás.[17] En la vida social y política, ésta parece ser la regla y no la excepción. El conjunto de la ciudadanía prefiere tal vez alguna constitución política (dentro de cierta gama de regímenes posibles) a ninguna en absoluto, porque la estabilidad a largo plazo es importante para permitirles planificar. Cuando el derecho está consolidado y es difícil de modificar, uno puede regular su comportamiento de conformidad con él. No obstante, cada grupo de interés quizá prefiera una constitución específica de la gama de las otras: los acreedores cabildean para que en la constitución se proscriba el papel moneda, cada partido político promueve el sistema electoral que más lo favorece, los que tienen un candidato fuerte a la presidencia quieren que el cargo tenga mucho poder, etcétera.

Los equilibrios de coordinación múltiples también aparecen cuando diferentes sociedades crean en un principio distintos patrones de peso,

longitud y volumen y luego descubren los beneficios potenciales de una solución común. Europa continental y el mundo anglosajón mantienen criterios diferentes en esta materia. A diferencia del caso de las diversas soluciones constitucionales, el obstáculo al acuerdo no es una divergencia permanente de intereses, sino los costes de corto plazo de la transición. La elección del patrón también podría ser, sin embargo, un juego de la gallina. Supongamos, de manera poco verosímil, que el patrón está incorporado a la constitución como una cláusula inamovible (inmune a las enmiendas). Cada país tendrá entonces un incentivo para comprometerse antes de que el otro lo haga.

El juego del teléfono se define por la necesidad de una regla que indique a las partes qué hacer cuando una conversación telefónica se interrumpe accidentalmente. Hay dos equilibrios de coordinación: la nueva llamada es hecha por la persona que ha telefoneado en primer lugar o por la persona receptora. Cualquiera de las dos soluciones es mejor que hacer que ambas vuelvan a llamar o que ninguna lo haga. Sin embargo, en este caso, a diferencia de la batalla de los sexos, un equilibrio es mejor para ambas que el otro. Lo más eficiente es que vuelva a llamar el emisor, ya que cuenta con más probabilidades de saber a qué número hacerlo. Los agentes racionales y plenamente informados convergirán en el equilibrio de coordinación superior. Si el coste es grande, el juego se convierte en una batalla de los sexos.

Consideremos, para terminar, el juego del punto focal, que puede ilustrarse con una variante de la batalla de los sexos. Los esposos han acordado ver una película que exhiben en los cines A y B, pero han postergado la elección del lugar. Suponemos que ninguno de los cines está más cerca o es en otros aspectos más conveniente que el otro. Como en la batalla de los sexos, la información, la racionalidad y el conocimiento general no les dirán por sí mismos dónde ir. Sin embargo, en la situación podría haber una señal psicológica que sirva como «punto focal» para la coordinación. Si la pareja tuvo su primera cita en el cine A, este hecho podría hacerlos coincidir en ese lugar. En este caso, la señal es un suceso puramente privado. En otros casos, las señales podrían ser compartidas por una amplia población. Entre los neoyorquinos, por ejemplo, el folclore dice que si dos personas quedan separadas por alguna causa, deben reunirse al mediodía bajo el reloj principal de la estación Grand Central. Y aun cuando no haya folclore, muchas personas irían a la estación de trenes, dado que en

muchas ciudades ésta es, entre los edificios de los que sólo hay uno por categoría, el más importante.[18] Su unicidad la hace atractiva como punto focal. El mediodía tiene la misma propiedad.[19]

Este punto focal es fácil de demostrar en los experimentos. Si se pide a todos los integrantes de un grupo que escriban un número entero positivo en un pedazo de papel y se les dice que recibirán una recompensa en caso de escribir todos el mismo, convergirán invariablemente en el 1. El entero más pequeño es uno solo, pero no hay un único entero más grande. En otros contextos, el cero puede aparecer como el único punto focal. Durante la guerra fría, cuando se debatía si los Estados Unidos podían utilizar armas nucleares tácticas sin desatar una escalada que culminara en la guerra nuclear total, se sugerían varias ideas acerca de una «línea roja» que autorizara un uso limitado. En definitiva, se decidió que el único punto focal consistía en *no usarlas*.

Pascal hizo una observación similar sobre la importancia de la costumbre: «¿Por qué se siguen las antiguas leyes y las antiguas opiniones? ¿Es que son las más sanas? No, sino que son las únicas, y nos cercena la raíz de la diversidad». En otro lugar escribió:

> Las cosas del mundo más fuera de razón se convierten en las más razonables a causa del desarreglo de los hombres. ¿Qué hay menos razonable que elegir para gobernar un Estado el primogénito de una reina? No se elige para gobernar una nave al viajero que es de mejor casa: esa ley sería ridícula e injusta. Pero como lo es, y lo será siempre, viene a ser razonable y justa; porque ¿a quién se elegiría? ¿Al más virtuoso y más hábil? Henos ahí *incontinenti* irnos a las manos: cada uno pretende ser el más virtuoso y más hábil. Atribuyamos, por consiguiente, esa cualidad a alguna cosa incontestable. Es el hijo primogénito del rey; esto es evidente, no cabe disputa. La razón no puede ser mejor, porque la guerra civil es el mayor de los males.

En realidad, este razonamiento puede influir en la elección de un rey cuando hay varios pretendientes al trono. En la elección del monarca durante la Restauración francesa, Talleyrand argumentó con éxito que el heredero legítimo del último rey de Francia era el único punto focal que podía impedir conflictos divisionistas. Según escribió en sus *Memorias*: «Un rey *impuesto* sería el resultado de la fuerza o la intriga; en uno y otro caso, resultaría insuficiente. Para establecer un sistema duradero que sea aceptado sin oposición, es preciso actuar sobre la base de un principio». Más adelante, Marx sostuvo que la República de 1848 debía su existencia al hecho de que no era la mejor opción para ninguna de las dos ramas de la familia real. Tocqueville hizo una observación similar para explicar la estabilidad del régimen de Napoleón III. También la democracia puede

verse como una solución de punto focal. Cuando hay muchos terrenos cualitativos antagónicos en función de los cuales la gente puede reivindicar su superioridad (sabiduría, riqueza, virtud, nacimiento), la solución cuantitativa del gobierno de la mayoría cobra una preponderancia única. Los ex países coloniales donde las tribus hablan diferentes lenguas pueden elegir el idioma del colonizador con finalidades oficiales. Las partes litigantes coinciden con facilidad en una propuesta que no es la mejor opción para ninguna.

En 1989, el traslado de los restos de Imre Nagy proporcionó un punto focal para que doscientas cincuenta mil personas marcharan por las calles de Budapest con el propósito de mostrar su desafección con el régimen. Como en los ejemplos anteriores, ese punto focal permitió la cooperación a través de la coordinación. En situaciones conflictivas, los puntos focales pueden tener efectos muy diferentes. En la guerra de Crimea, el general francés Pélissier decidió lanzar el segundo ataque contra Sebastopol el 18 de junio de 1855, porque quería complacer a Napoleón III obteniendo una victoria en el aniversario de la batalla de Waterloo. Como esta fecha y su importancia para los franceses eran de conocimiento general, los rusos pudieron anticiparse y derrotarlo.

Una lección de este examen es que una situación dada del mundo real se puede modelar en el marco de varios juegos diferentes, según cuáles sean los supuestos adicionales. La carrera armamentística se modeló como DP, como juego de la gallina y como JS. La afiliación al sindicato puede ser un DP o un JS. La repetición de la llamada telefónica se ha visto como una batalla de los sexos o un juego del teléfono. La coordinación de pesos y medidas podría ser un juego de la gallina o una batalla de los sexos. El grano fino de las estructuras de interacción quizá no sea inmediatamente visible. Al obligarnos a ser explícitos acerca de la naturaleza de la interacción, la teoría de los juegos puede revelar sutilezas o perversidades insospechadas.

Juegos secuenciales

Me gustaría ocuparme más brevemente de los juegos en que los agentes toman *decisiones secuenciales* (analizo estos juegos con mayor detalle en el capítulo siguiente), y comenzar con un sencillo ejemplo demostrativo del

poder de la teoría de los juegos para esclarecer estructuras de interacción que antes sólo se comprendían de manera oscura.[20]

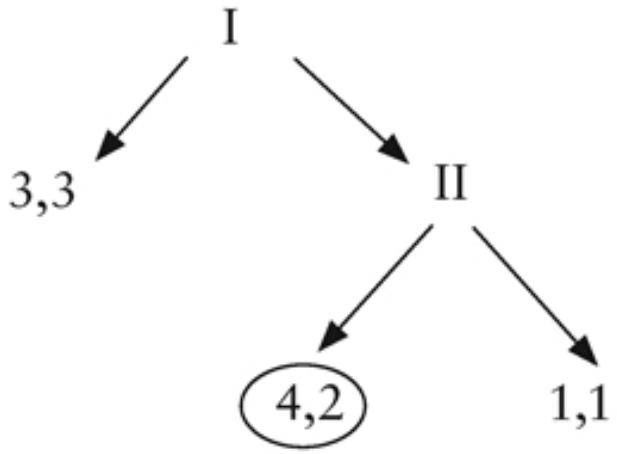


FIGURA 19.5

En la Figura 19.5, dos ejércitos se enfrentan en la frontera de sus países. El general I puede o bien retirarse, y mantener así el statu quo (3, 3), o invadir. Si invade, el general II puede o bien combatir, con el resultado (1, 1), o bien ceder un fragmento disputado del territorio, con el resultado (4, 2). Antes de que I tome su decisión, II tal vez pueda dar a conocer su intención de combatir en caso de ser atacado, con la esperanza de inducir a I a escoger (3, 3) y no (1, 1). Sin embargo, *esa amenaza no es creíble*. I sabe que una vez que invada, el interés de II será ceder y no combatir. El único resultado de equilibrio es (4, 2). Este concepto de equilibrio no es el concepto estático de la «mejor respuesta» que hemos analizado hasta aquí. Antes bien, se trata de un concepto dinámico que comienza con las últimas etapas del juego y se retrotrae a las primeras. (El término técnico es «retroinducción».) En primer lugar, nos preguntamos cuál sería la actitud racional de II si I invadiera. La respuesta: «ceder», conduce al resultado (4, 2). En consecuencia, I tiene que elegir entre un curso de acción que lleva a (3, 3) y otro que lleva a (4, 2). Como es racional, escoge el último.

Según Tucídides observó en la *Historia de la Guerra del Peloponeso*, las promesas también tienen que ser creíbles para que el otro lado las tome como base de su comportamiento.

> Los juramentos hechos en apoyo de una reconciliación sólo tenían momentánea validez, pues cada bando los hacía únicamente en ausencia de todo otro medio de fuerza para librarse de un callejón sin salida; mas quienquiera que encontrase al adversario desprevenido en un momento dado y aprovechara la oportunidad para dar un audaz golpe, disfrutaba, por haber explotado la buena fe, de una venganza más dulce que si hubiera vencido en abierto combate. […] Pues ninguna palabra era lo bastante confiable, y ningún juramento lo bastante imponente, para traer la reconciliación, y todos los que se encontraban en una posición superior, imaginando que la seguridad no podía siquiera esperarse, tomaban medidas para evitar el agravio y no para permitirse confiar en nadie.

Para decirlo de otro modo, la persona a quien se hacía la promesa debía preguntarse si para quien prometía era racional mantener su palabra. Si se permitiera la comunicación en el juego de la confianza (Capítulo 15), por ejemplo, el segundo jugador podría tratar de inducir al primero a hacer una gran transferencia, con la promesa de un gran reembolso. Si no hay nada que lo haga atenerse a su palabra, la promesa no es creíble. En *La democracia en América*, Tocqueville comenta con sarcasmo una carta del secretario de guerra a los cheroquis, en la cual «les declara que no deben vanagloriarse de conservar el disfrute del territorio que ocupan en ese momento, sino que les da la misma seguridad positiva para el tiempo en que se encuentren del otro lado del Misisipí […], *¡como si el poder que les falta ahora no debiera faltarles del mismo modo entonces!*» En China, la reforma económica ha sido vulnerable a un problema similar. Cuando el Gobierno estableció reformas de mercado en la agricultura, prometió a los agricultores arriendos de tierras por quince años con el fin de darles un incentivo para mejorarlas. Como no hay manera de hacer que un gobierno autocrático se atenga a sus promesas, muchos agricultores no le creyeron y destinaron las ganancias al consumo. Un gobierno autocrático es *incapaz de garantizar su incapacidad* de interferir.

La noción de credibilidad es central en la «segunda generación» de la teoría de los juegos, iniciada alrededor de 1975. (La primera generación se remonta aproximadamente a 1945.) Una vez que tomamos en serio la idea, nos vemos en la necesidad de preguntarnos cómo podrían los agentes *invertir en credibilidad* para prestar eficacia a sus amenazas y promesas. Los mecanismos son varios. Uno es la *construcción de reputación*: invertir, por ejemplo, para granjearse la fama de ser un tanto u ocasionalmente irracional. Así, se ha dicho que, alentado por Henry Kissinger, el presidente Nixon cultivó en forma deliberada un estilo errático para hacer creer a los soviéticos que, si lo provocaban, podía llegar a actuar contra el interés estadounidense. Por otra parte, la gente podría llevar a la práctica sus

amenazas, aunque no redundara en su interés hacerlo, con el objeto de construirse una reputación de dureza que impulsara a otros a creer en sus amenazas en oportunidades ulteriores.

Otro mecanismo es el *precompromiso*, examinado en el Capítulo 13. Allí, se lo veía como una respuesta racional secundaria a la proclividad del agente a comportarse irracionalmente. En el contexto estratégico, el precompromiso puede ser plenamente racional. En el juego representado en la Figura 19.5, el general II podría construir una «máquina del Juicio Final» que lanzara automáticamente un ataque nuclear contra el otro país en caso de invasión. Si tanto la existencia de esa máquina como el hecho de que su funcionamiento escapa al control del país II son de conocimiento general, la situación actuaría como un disuasivo de la invasión. O bien II podría valerse de la estrategia consistente en «quemar los puentes», es decir, cortar toda posibilidad de retirada. También entonces el general I dejaría a un lado sus propósitos si supiera que el general II no tiene otra opción que combatir la invasión.

En algunos casos, ambas partes pueden tratar de usar el precompromiso para obtener una ventaja sobre la otra. En las negociaciones entre los trabajadores y la patronal, las amenazas de huelgas y boicots tal vez no sean creíbles. La patronal sabe que, como los trabajadores tienen hipotecas que pagar y familias que alimentar, no pueden darse el lujo de hacer una huelga muy larga. El sindicato sabe que, como la empresa tiene que cumplir contratos, no puede darse el lujo de interrumpir la producción. Para fortalecer la credibilidad de sus amenazas, el sindicato podría invertir en un fondo de huelga (tal vez en conjunto con otros sindicatos), y la patronal, en la acumulación de grandes existencias. O bien los negociadores de cada una de las partes podrían anunciar públicamente sus exigencias mínimas y sus ofertas máximas, para asegurarse de que incurrirán en elevados costes de reputación si ceden. Ese «juego de precompromiso» podría ser o bien un DP o bien un juego de la gallina, según cuál fuera la estructura del juego subsiguiente.

* * *

Nota bibliográfica

La mejor introducción elemental a la teoría de los juegos es Avinash Dixit y Susan Skeath, *Games of Strategy*, segunda edición, Nueva York,

Norton, 2004. Entre otros tratamientos más avanzados, sugiero Fernando Vega-Redondo, *Economics and the Theory of Games*, Cambridge y Nueva York, Cambridge University Press, 2003. Una exploración enciclopédica con muchas aplicaciones es Robert J. Aumann y Sergiu Hart, *Handbook of Game Theory with Economic Applications*, tres volúmenes, Ámsterdam, North-Holland, 1992, 1994 y 2002. Se encontrarán aplicaciones a tópicos específicos en James D. Morrow, *Game Theory for Political Scientists*, Princeton, Princeton University Press, 1994, y Douglas G. Baird, Robert H. Gertner y Randal C. Picker, *Game Theory and the Law*, Cambridge (Massachusetts), Harvard University Press, 1994. Un estudio clásico de las convenciones es David K. Lewis, *Convention: A Philosophical Study*, Cambridge (Massachusetts), Harvard University Press, 1969, que se inspira ampliamente en otro clásico, Thomas C. Schelling, *The Strategy of Conflict*, Cambridge (Massachusetts), Harvard University Press, 1960 [*La estrategia del conflicto*, Madrid, Tecnos, 1964], en el cual se expuso por primera vez la idea de los puntos focales. La obra de Schelling también proporcionó el fundamento intuitivo a la «segunda generación» de la teoría de los juegos, formalmente elaborada por Reinhard Selten, «Reexamination of the perfectness concept for equilibrium points in extensive games», *International Journal of Game Theory*, 4(1), 1975, págs. 25-55. Se encontrarán varias técnicas de precompromiso para los juegos políticos en James Fearon, «Domestic political audiences and the escalation of international disputes», *American Political Science Review*, 88(3), septiembre de 1994, págs. 577-592. Para su uso en la negociación salarial, véase mi *The Cement of Society: A Study of Social Order*, Cambridge y Nueva York, Cambridge University Press, 1989 [*El cemento de la sociedad: las paradojas del orden social*, Barcelona, Gedisa, 1997].

Notas al pie

[7] Un hecho es de conocimiento general si todos lo saben, todos saben que todos los otros lo saben, todos saben que todos los otros saben que todos los otros lo saben, y así sucesivamente. Para evitar apoyarse en la expresión «y así sucesivamente», que sugiere una secuencia infinita de creencias, la idea también puede enunciarse del siguiente modo: no hay *n* tal que el hecho sea de conocimiento general hasta el nivel *n* en la secuencia, pero no en el nivel *n* + 1. Una ilustración simple es el conocimiento general que puede alcanzarse en un aula. Cuando el docente cuenta un hecho a los alumnos, todos lo saben, saben que los otros lo saben, y así sucesivamente.

[8] Por convención, el primer número de cada celda representa la retribución para el «jugador de las hileras» que elige entre la estrategia de arriba y la de abajo, y el segundo, la retribución para el «jugador de las columnas» que elige entre la estrategia izquierda y la derecha. Según cuál sea el contexto, las retribuciones pueden ser utilidades cardinales, utilidades ordinales, dinero o cualquier otra cosa que pueda suponerse maximizada por los jugadores. En la Figura 19.1, puede considerarse que las retribuciones representan utilidades ordinales, como un reflejo de las preferencias sobre los resultados. Aquí y más adelante, los equilibrios están rodeados por un círculo.

[9] De acuerdo con algunos análisis jurídicos, una función importante del derecho de la responsabilidad extracontractual [*tort law*] es utilizar el sistema de multas e indemnizaciones para modificar la matriz de recompensas, de modo que el equilibrio emergente tenga alguna propiedad deseable (eficiencia o equidad).

[10] Aunque la no unicidad no se deduce de la definición formal, éste parece ser un rasgo general de los juegos de coordinación de la vida real.

[11] Aunque expuestas aquí como juegos de dos personas, es fácil generalizarlas al caso de muchos agentes. La Figura 24.2, por ejemplo, ilustra una versión de *n* + 1 personas del dilema del prisionero.

[12] Las retribuciones para el dilema del prisionero de la Figura 19.3 acaso parezcan artificiales. Sin embargo, a los efectos presentes, lo único que importa es la clasificación (ordinal) de los resultados. Más adelante, las retribuciones se reinterpretarán como recompensas monetarias.

[13] «Si quieren cazar un ciervo, todos [los cazadores] advierten que, para lograrlo, deben mantenerse sin vacilaciones en su lugar: pero si una liebre acertara a ponerse al alcance de cualquiera de ellos, éste, a no dudar, habría de perseguirla sin parar en mientes y, tras cobrar su presa, muy poco lo inquietaría que, al hacerlo, llevara a los otros a perder las suyas.» Una interpretación del párrafo podría ser que, para Rousseau, la persecución de liebres es una estrategia dominante.

[14] Ésta es una enorme simplificación, planteada con meros fines de ilustración.

[15] Ya utilicemos castigos o recompensas, los costes de establecer el sistema y supervisar a los agentes también tienen que financiarse mediante las ganancias de la cooperación. En la práctica, es muy fácil que esta situación convierta esos arreglos en imposibles o dispendiosos.

[16] De manera similar, la «solución» al dilema del prisionero consistente en que cada persona prometa cooperar no hace sino recrear el juego, ahora con las opciones de «mantener la promesa» y «renegar».

[17] Como veremos más adelante (Capítulo 25), esta cuestión de dividir los beneficios de la cooperación también puede estudiarse en el marco de la *teoría de la negociación*, una rama más especializada de la teoría de los juegos.

[18] En Nueva York, quienes desconozcan el folclore no irán a Grand Central, dado que la existencia de la estación Penn hace que no sea única. Podrían, en cambio, coordinar un encuentro en el Empire State Building.

[19] Si bien la medianoche también es un punto focal, es inferior al mediodía debido a la inconveniencia.

[20] Mantengo el supuesto de que la racionalidad y la información son de conocimiento general.

Capítulo 20
Juegos y comportamiento

Intenciones y consecuencias

La estructura conceptual de la teoría de los juegos es esclarecedora. ¿Nos ayuda también a *explicar el comportamiento*? Consideremos la justificación para quemar los puentes o las naves basada en esa teoría. Este comportamiento podría adoptarse por las razones estratégicas expuestas en el capítulo anterior, pero también por otras. El *Oxford English Dictionary* presenta la siguiente cita de *Tarzán de los monos*, de Edgar Rice Burroughs: «Como ella había tenido miedo de sucumbir a las súplicas de ese gigante, había quemado los puentes tras de sí». No se trata de un ejemplo de razonamiento estratégico. Antes bien, la mujer en cuestión parece temer la posibilidad de ceder a los ruegos si no hace lo imposible para impedírselo. Aun en la esfera militar, esas justificaciones no estratégicas podrían ser tan importantes como las estratégicas. Un comandante tal vez queme los puentes para que el miedo a sus enemigos no impulse a sus soldados a huir. Quizá quiera impedir su *propia* deserción, si teme rendirse a la debilidad de la voluntad. El comandante A podría quemar los puentes o las naves para señalar a B, el comandante enemigo, que no puede contar con la huida de las tropas de A. Ése fue al parecer el razonamiento de Cortés cuando, tras decir a su tripulación (de manera creíble pero no veraz) que su flota no podía hacerse al mar, quemó todos sus barcos salvo uno. (Además, al quemar las naves podía incorporar a los marineros a su infantería.) Para diferenciar entre estas diversas explicaciones, es preciso determinar las

intenciones de los agentes. Los beneficios concretos de la quema de los puentes no son ni necesarios ni suficientes para establecer una explicación en términos de los beneficios esperados (Capítulo 3).

Aunque la teoría de los juegos explica el comportamiento apelando a las intenciones de los actores de provocar ciertas consecuencias, también puede dar razón de situaciones en las que éstas tienen sin cuidado a algunos de los actores. Consideremos, por ejemplo, la interacción entre la Unión Europea y los nuevos ingresantes de Europa oriental. Los viejos Estados miembros podrían estar tentados de imponer condiciones de ingreso que implicaran subsidios agrícolas permanentemente inferiores para los nuevos Estados, en comparación con los antiguos miembros de la misma superficie territorial. En términos materiales, los nuevos Estados obtendrían tantos beneficios con el ingreso que estarían mejor como miembros de segunda que como no miembros, aunque no tan bien como si fueran miembros plenos. En términos psicológicos, el insulto que significa ser tratado como inferior podría llevarlos a rechazar tales condiciones.[21] En previsión de esa reacción, los viejos Estados podrían proponer el ingreso en condiciones de total igualdad. La creencia de que las condiciones materiales no son la preocupación exclusiva de los nuevos Estados podría ayudarles a conseguir mejores condiciones.

Como no he tenido conocimiento particular de las negociaciones de ingreso, estas observaciones son meras conjeturas. Sabemos, sin embargo, que argumentos de ese tipo se plantearon en la Convención Federal de Filadelfia en 1787, durante el debate por las condiciones de ingreso de los futuros estados del oeste. El gobernador Morris y otros proponían que éstos fueran admitidos como estados de segunda categoría, con el fin de que nunca pudieran imponerse por el voto a los trece estados originales. Contra este punto de vista, George Mason sostuvo con vigor que era preciso incorporarlos con los mismos derechos que los estados fundadores de la Unión. En primer lugar, su argumento apelaba a los principios: al admitir a los estados del oeste en iguales condiciones, los redactores harían «lo que sabemos justo en sí mismo». Y Mason agregaba, para quienes no aceptaran el argumento, que, de todos modos, era improbable que los nuevos estados dieran su consentimiento a una propuesta degradante.

> Si los estados occidentales deben ser admitidos en la Unión a medida que se establezcan, será menester tratarlos como iguales y no someterlos a discriminaciones degradantes. Tendrán el mismo orgullo y las demás pasiones que nos mueven, y no se incorporarán a la Unión o se

> rebelarán rápidamente contra ella si no se los pone en todos los aspectos en un pie de igualdad con sus hermanos.

Mason se refiere al «orgullo y las pasiones» de los nuevos estados, no a su propio interés. Aun cuando les interesara en efecto acceder a la Unión en condiciones desiguales en vez de permanecer al margen, podrían, por resentimiento, preferir mantenerse fuera. Al mismo tiempo, Mason apela al egoísmo de los viejos estados, no a su sentido de justicia. En la terminología del Capítulo 4, les dice que, como la motivación de los nuevos estados podría ser *la pasión y no el interés*, redundaría en el interés de los viejos estados actuar como si los motivara *la razón y no el interés*.

Esta situación ha sido objeto de estudios experimentales por medio de los juegos del ultimátum y el dictador (Figura 20.1). En el primero de estos juegos, una persona (el proponente) puede proponer una distribución (x, 10 $–x$) de diez dólares entre él mismo y otra persona (el copartícipe). Los ofrecimientos sólo pueden hacerse en dólares enteros. Si el copartícipe acepta, esa distribución se lleva a la práctica. Si el copartícipe rechaza la propuesta, ninguno de los dos obtiene nada. Aunque el juego se ha estudiado en muchas variantes, me concentro en interacciones únicas en condiciones de anonimato. Como los sujetos interactúan a través de terminales de ordenador, no conocen la identidad de su compañero. A menudo se les aclara que el experimentador no podrá determinar quién ha hecho qué elección, con lo cual se elimina la posibilidad de que sus decisiones estén influenciadas por el deseo de complacerlo. Cuando los sujetos juegan muchas veces, nunca tienen el mismo compañero; esta situación permite el aprendizaje, pero no la construcción de una reputación. En esas condiciones, las decisiones tienen un margen máximo para reflejar un egoísmo irrestricto.

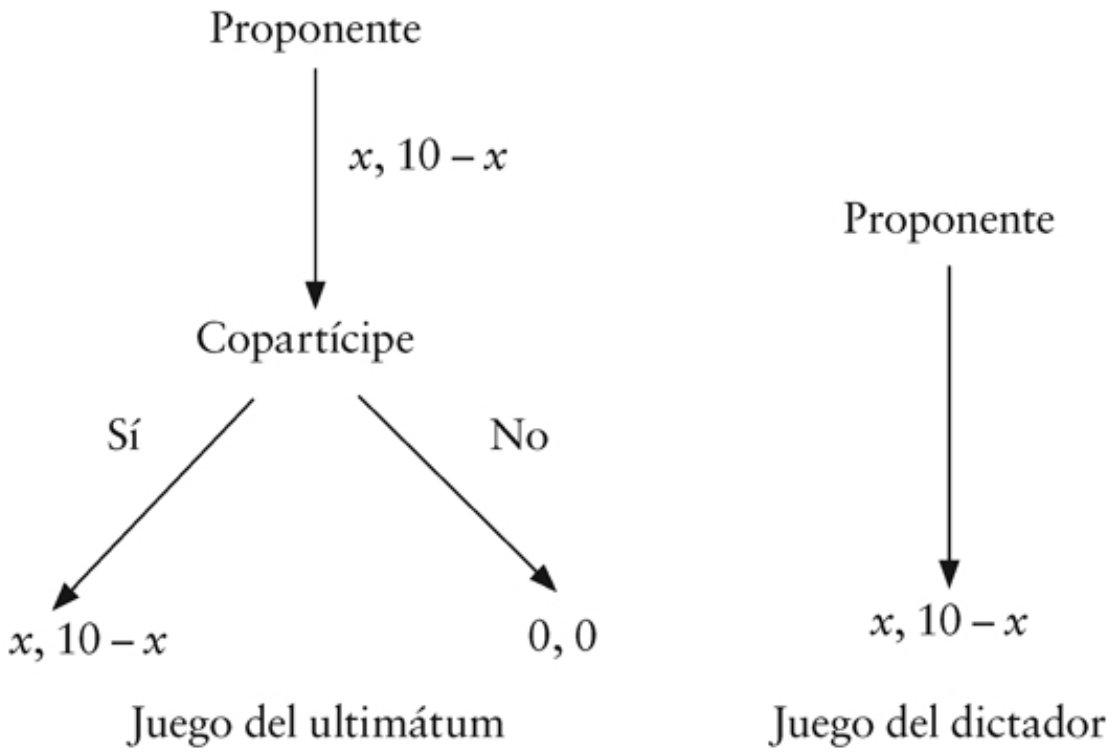


FIGURA 20.1

Suponiendo que ambos agentes son racionales y egoístas y tienen plena información sobre la estructura de retribuciones, y que estos hechos son de conocimiento general, el proponente ofrecerá (9, 1), que el copartícipe aceptará. Si los ofrecimientos pudiesen hacerse con centavos, el proponente ofrecería (9,99, 0,01) y aun así la cifra se aceptaría, porque algo es mejor que nada. En los experimentos, las propuestas están habitualmente en torno de los (6, 4). Los copartícipes suelen rechazar las propuestas que les ofrecen dos o menos.[22] Están dispuestos a escupir al cielo. Es evidente que uno de los supuestos queda invalidado. En virtud de la configuración del experimento, podemos excluir la falta de información, así como la falta de conocimiento general sobre la información. No podemos excluir, en cambio, un fracaso de la racionalidad o la presencia de motivaciones no egoístas.

El proponente podría ser un *altruista* que prefiere una asignación más o menos igual a otra en la que él se lleva todo. Aunque el altruismo con perfectos extraños que no se encuentran en una necesidad evidente es una idea un tanto rara, la actitud es al menos compatible con la racionalidad. Sin

embargo, podemos rechazar esta hipótesis si comparamos el comportamiento en el juego del ultimátum con el comportamiento en el juego del dictador. En este último, que en realidad no es en absoluto un «juego», el proponente asigna unilateralmente el dinero entre él mismo y el copartícipe, quien, de tal modo, no tiene oportunidad alguna de responder. Si en el juego del ultimátum las propuestas sólo fueran dictadas por el altruismo, las asignaciones en el juego del dictador no deberían ser diferentes. En los experimentos, no obstante, el comportamiento del proponente es mucho menos generoso en este último juego. No parece haber dudas de que en el juego del ultimátum, la expectativa de rechazo de los ofrecimientos no generosos moviliza, al menos en parte, el comportamiento del proponente.

Para explicar ese rechazo, podríamos suponer que la *envidia* motivará a los copartícipes a rechazar los ofrecimientos pobres, y que los proponentes egoístas, al prever ese efecto, harán ofrecimientos cuya generosidad será apenas la suficiente para ser aceptados. Si esta explicación es correcta, cabe esperar que la frecuencia de rechazo de (8, 2) sea la misma cuando el proponente tiene la libertad de proponer cualquier reparto y cuando está obligado –y sabe que lo está– a elegir entre (8, 2) y (2, 8). En los experimentos, el índice de rechazos es más bajo en el último caso. Este resultado sugiere que el comportamiento del copartícipe está determinado por consideraciones de *equidad*. El ofrecimiento del proponente de (8, 2), cuando podría haber ofrecido (5, 5), se juzga menos equitativo que cuando su única alternativa es una propuesta igualmente desventajosa para él. Lo que cuenta son las *intenciones*, no los resultados.

Esta interpretación es respaldada por la importancia de la reciprocidad fuerte en otros juegos, como el de la confianza (Capítulo 15). En ocasiones, la gente está dispuesta a castigar a otros por comportarse injustamente, a pesar de que eso signifique algún coste y ningún beneficio para sí misma. Esta práctica parece transgredir uno de los cánones de la racionalidad enumerados en el Capítulo 12: en una elección entre actuar y no hacer nada, un agente racional no actuará si los costes esperados son superiores a los beneficios esperados. Las explicaciones en términos de altruismo o envidia no violarían ese principio. Para un altruista, el resultado puede ser mejor cuando beneficia a otro con algún coste para sí mismo, y para el envidioso, cuando perjudica a otro con algún coste para sí mismo. Ese comportamiento contraviene el supuesto del egoísmo, pero no el de la racionalidad. En

contraste, la explicación basada en la equidad parece infringir este último supuesto. La reciprocidad fuerte induce un comportamiento similar al que mostramos cuando tropezamos con una piedra y, como represalia, le damos un puntapié: no ayuda y no hace sino agravar el dolor.[23]

Retroinducción

En el juego del ultimátum, el juego mostrado en la Figura 19.5 y otros juegos secuenciales, el equilibrio se encuentra mediante la retroinducción. En el primero de los juegos mencionados, el proponente prevé la respuesta del copartícipe a una propuesta determinada y luego hace el ajuste correspondiente en su conducta. En estos ejemplos, los cálculos necesarios son muy simples. En otros experimentos, los sujetos podrían tener que embarcarse en cadenas más largas de razonamientos. Es posible, por ejemplo, que se diga a dos sujetos que pasen por tres rondas de ofertas y contraofertas para dividir una suma de dinero, que se reduce un 50% luego de cada ronda.[24] En cada momento, un agente puede o bien aceptar la propuesta e ir hacia la «derecha» o hacer una contrapropuesta e ir hacia «abajo». La racionalidad, el egoísmo y el conocimiento general inducen luego el siguiente razonamiento.

La persona que hace la primera propuesta (jugador I) deberá tomar en cuenta la posibilidad de que el jugador II prefiera esa división a otra en la que recibe una porción más grande de un pastel más pequeño. Al mismo tiempo, el jugador I sabe que el jugador II no hará una propuesta que, de ser aceptada por I, ponga a éste en peor situación de lo que lo estará si llega hasta la última ronda. En la Figura 20.2, el jugador I puede al menos obtener 1,25 si toma todo lo que queda en la tercera ronda. Por lo tanto, el jugador II no puede ofrecerle menos de 1,25 en la segunda, dejando esa cifra como máximo para sí mismo. Al saber esto, el jugador I ofrecerá (3,75, 1,25) y II la aceptará.

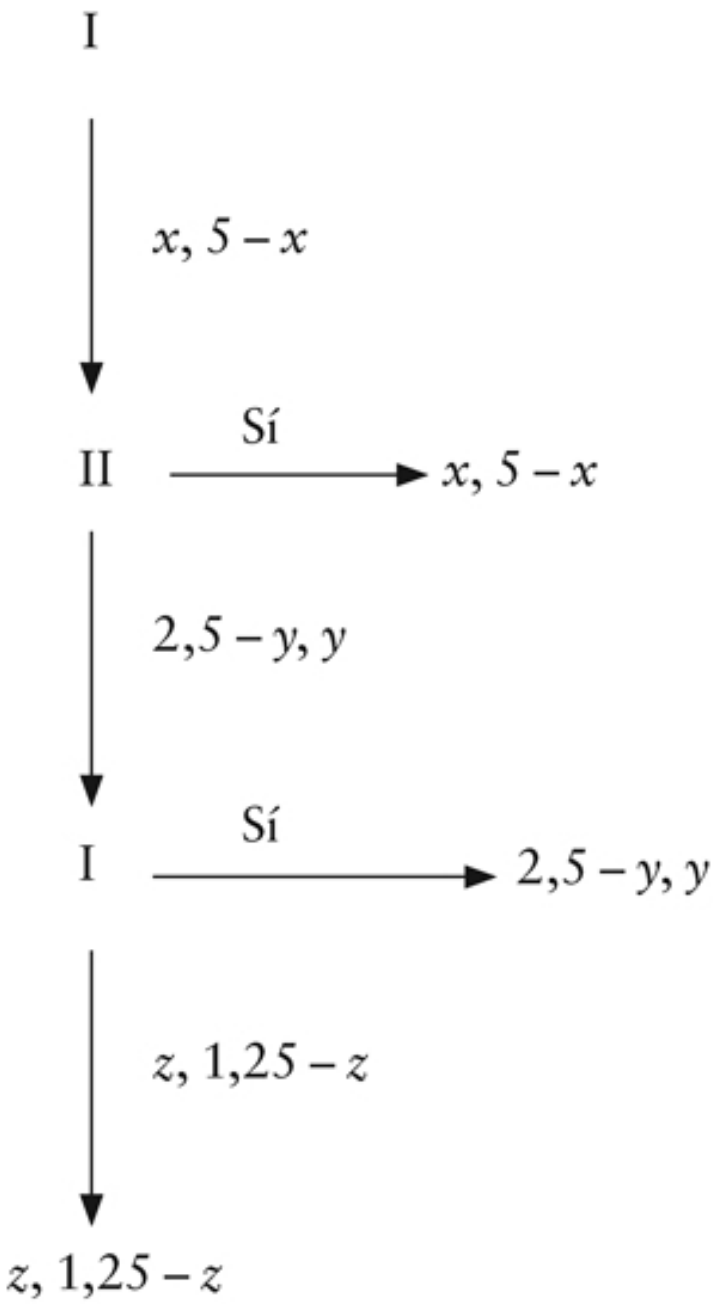


FIGURA 20.2

En los experimentos, el ofrecimiento medio hecho por I es (2,89,2,11), sustancialmente más generoso que la oferta de equilibrio. Es evidente que se contravienen uno o más supuestos. 1) El primer jugador podía ser altruista. 2) Podría temer que el otro jugador rechazara la oferta de equilibrio porque es incapaz de seguir la lógica de la retroinducción. 3) Él

mismo podría ser incapaz de seguir esa lógica.[25] Para eliminar la primera y la segunda hipótesis, se puede decir a los sujetos que están jugando contra un ordenador programado para dar respuestas óptimas. En ese caso, el ofrecimiento medio es de (3,16, 1,84), que sigue siendo sustancialmente más alto que el equilibrio. Como los sujetos que hacen los ofrecimientos elevados difícilmente puedan tener sentimientos altruistas hacia un ordenador, o creer que éste es incompetente, ellos mismos deben ser incompetentes.

No es que la tarea sea ardua. Una vez que la lógica de la retroinducción se explica a los sujetos, éstos tienen un desempeño impecable en nuevos juegos. El experimento muestra, antes bien, que este tipo de razonamiento no se produce naturalmente en los seres humanos. Es posible que ni siquiera un simple razonamiento prospectivo ocurra de manera espontánea, como lo muestra la maldición del ganador (Capítulo 12). El síndrome del «hermano menor» (Capítulo 18) tiene un sabor parecido. No es que las personas, tras pensarlo bien, no puedan entender que los otros son tan racionales y capaces de deliberación como ellas mismas: sucede que tienden de manera espontánea a pensar que están aferrados a sus hábitos en vez de ver que se ajustan a sus entornos.

Algunos fracasos de la teoría de los juegos basada en la elección racional

Entre muchos otros descubrimientos que revelan fracasos predictivos de la teoría de los juegos, examinaré el «dilema del prisionero repetido una cantidad finita de veces», la «paradoja de la cadena de tiendas», el «juego del ciempiés», el «dilema del viajero» y el «concurso de belleza».

Cuando los sujetos juegan muchos DP sucesivos entre sí y saben cuál será la última ronda, observamos una proporción sustancial de elecciones de C, que a menudo superan el 30%. Una explicación intuitiva es que un jugador puede elegir C en una ronda con la esperanza de que el otro haga lo propio («hoy por ti, mañana por mí»). No obstante, si los jugadores adoptan la retroinducción, comprenderán que en el juego final ambos escogerán D, dado que no habrá oportunidades de influir sobre el comportamiento en un juego posterior. En el penúltimo juego también elegirán D, pues el comportamiento en el juego final está dado por el argumento previo. Este

argumento recorre «zumbando» todo el camino de vuelta hasta la primera ronda, y provoca así la defección en todos los juegos.

Una cadena de tiendas tiene sucursales en veinte ciudades, y en cada una de ellas enfrenta a un rival potencial. El rival tiene que decidirse entre instalar una tienda para compartir el mercado con esa cadena o mantenerse fuera de la ciudad. La cadena de tiendas tiene la opción de responder agresivamente mediante una rebaja feroz de los precios y, de tal modo, mandar a la ruina al competidor pero también autoimponerse una pérdida, o aceptar compartir el mercado. Las retribuciones se muestran en la Figura 20.3; el primer número de cada par es la retribución para el ingresante potencial.

La retroinducción en un solo juego da (5, 5) como resultado de equilibrio: el rival ingresa y la cadena acepta compartir el mercado. No obstante, en previsión de desafíos ulteriores, la cadena de tiendas podría decidir comportarse más agresivamente y arruinar al ingresante, con cierto coste para sí misma, a fin de disuadir a potenciales competidores en otras ciudades. Pero si aplicamos la retroinducción a la secuencia de veinte juegos, esta estrategia no es viable. En el vigésimo juego el comportamiento agresivo ya no redundará en más beneficios, de modo que la empresa también podría compartir el mercado con el ingresante. Sin embargo, esto implica que tampoco se obtengan beneficios con la política de precios rapaz en el decimonoveno juego, y así sucesivamente hasta llegar al primero. Aunque la magnitud de las políticas rapaces de precios en los mercados reales es un tema controvertido, sí aparece en los mercados experimentales.

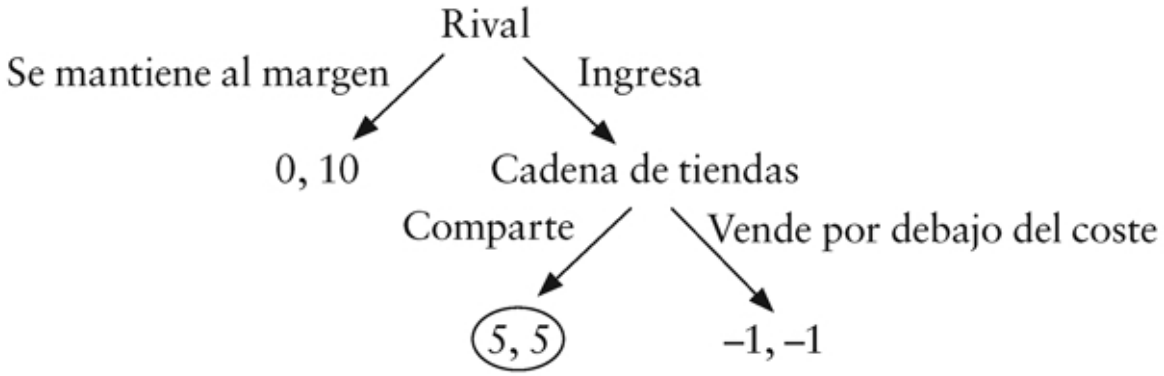


FIGURA 20.3

El juego del ciempiés[26] se muestra en la Figura 20.4 (retribuciones en dólares). La retroinducción indica al jugador I la elección de detenerse al comienzo, lo cual deja a cada uno de los dos participantes con una dieciseisava parte de las retribuciones que podrían haber obtenido de continuar hasta el final. En un experimento típico, el 22% escogió detenerse en el primer «nodo» de elección; el 41% de quienes permanecieron decidieron detenerse en el segundo nodo; el 74% de quienes aún se mantenían en el juego escogieron luego detenerse en el tercero, y de los que quedaban, la mitad eligió detenerse en el cuarto nodo y la mitad decidió seguir. La desviación con respecto al equilibrio (rodeado por un círculo) pronosticado por la retroinducción es amplia, al igual que el aumento promedio de las ganancias de los jugadores.

Para explicar estos ejemplos de cooperación y depredación aparentemente irracionales, podría estipularse la existencia de *incertidumbre* acerca de algunos aspectos de los juegos. En la vida real, los jugadores se ven en contadas ocasiones frente a un número finito y conocido de rondas. Es probable que muchas veces crean que la interacción va a continuar durante un tiempo indefinido, de modo que no hay una ronda final desde la cual pueda iniciarse la retroinducción. En tales casos, la adopción mutua de una actitud del tipo de «hoy por ti, mañana por mí» puede constituir un equilibrio en el DP reiterado. (No es único, dado que «siempre desertar, siempre desertar» también es un equilibrio. Estructuralmente, la situación se parece un poco al juego de la seguridad, con un buen equilibrio y uno malo.) Si la vida real induce un comportamiento de «uno por otro», los agentes pueden aplicarlo a situaciones de laboratorio en las cuales no es óptimo.

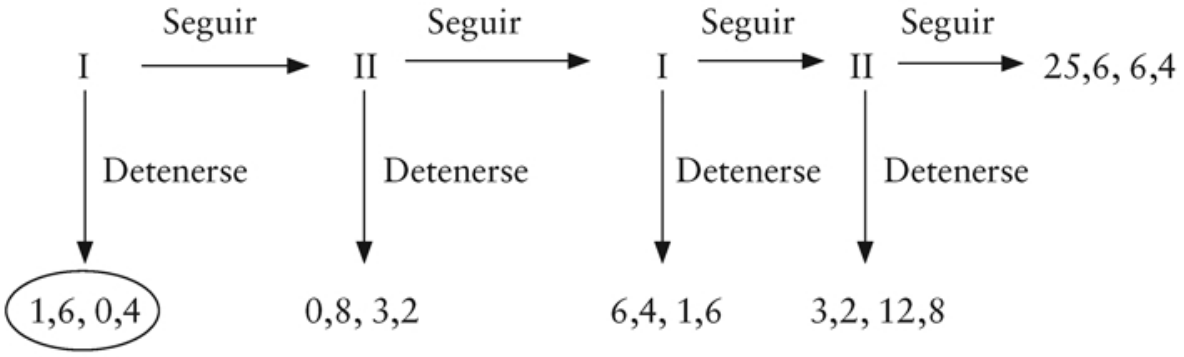


FIGURA 20.4

Como alternativa, un agente podría carecer de certeza con respecto al tipo de jugador a quien enfrenta. Supongamos que es de conocimiento general que en la población hay algunos individuos irracionales. Se sabe que algunos agentes siempre cooperarán, que otros utilizan la actitud de «hoy por ti, mañana por mí» en DP repetidos de manera finita, que otros se valdrán de una política rapaz de precios para disuadir a los ingresantes incluso en la vigésima ciudad, etc. Sin embargo, no se sabe con exactitud quiénes son esos individuos. Hay cierta probabilidad positiva de que cualquier agente sea irracional. En la paradoja de la cadena de tiendas, un ingresante potencial cambiará de opinión si atribuye una probabilidad bastante grande al carácter irracional del gerente de la tienda. Sabiéndolo, el gerente tiene un incentivo para embarcarse en una política rapaz de precios con el primer ingresante, para hacer creer a otros que es irracional. Cuando los ingresantes potenciales en otras ciudades observen esa situación, utilizarán un razonamiento bayesiano (Capítulo 11) para asignar una probabilidad más alta a la irracionalidad del gerente. Tal vez no sea lo suficientemente alta para disuadirlos, pero si el gerente lo hace una y otra vez, a la larga podrá alcanzarse un nivel en que para los rivales sea más racional no ingresar. Un argumento similar podría explicar la cooperación en el DP reiterado de manera finita y el juego del ciempiés.

Otra posibilidad es que en el dilema del prisionero repetido y en el juego del ciempiés la cooperación tenga en parte un carácter de punto focal. Aunque los individuos *racionales* desertaran en la primera ocasión, las personas *razonables* no lo harían. Si bien esta sugerencia (sobre la cual abundaré más adelante) es bastante vaga, parece más cierta, al menos a mi entender, que los argumentos basados en la incertidumbre acerca de la índole de la otra persona. Por un lado, esos argumentos exigen que los jugadores realicen cálculos enormemente complicados, que ocupan muchas páginas en los manuales. Por otro, la introspección y la observación casual indican que, al tomar decisiones en la vida diaria, no vemos de ese modo a los otros. Cuando confío a alguien una pequeña suma de dinero, pero no una grande, no lo hago por asignar una escasa probabilidad a que sea incondicionalmente confiable, sino porque juzgo que sólo se puede confiar en él cuando lo que está en juego no es de mucha magnitud.

En el dilema del viajero, dos jugadores presentan al mismo tiempo reclamos, por un valor de entre ochenta y doscientos dólares, por equipajes perdidos. Para desalentar los pedidos exorbitantes, la línea aérea paga a cada viajero el mínimo de los dos reclamos, añade una suma R para la persona que ha hecho el menor y deduce el mismo monto a quien ha hecho el reclamo más alto. Consideremos un par de reclamos de (100, 150), que otorgan retribuciones de (100 + R, 100 – R). Este par no puede ser un equilibrio, dado que el primer jugador tendría un incentivo para reclamar 149, con una retribución de 149 + R, a lo cual el segundo respondería con un reclamo de 148, y así sucesivamente. Como lo sugiere este ejemplo, el equilibrio único se alcanza cuando ambos reclaman 80. En experimentos, ese resultado se observa, de hecho, cuando R es grande. Sin embargo, cuando es pequeño, los sujetos hacen reclamos más cercanos al límite superior de 200. Una vez más, mi intuición es que actúa algo parecido al razonamiento del punto focal. Cada viajero sabe que, vistas las ganancias que pueden obtenerse al coordinar un reclamo elevado, sería tonto adoptar la estrategia de equilibrio, y espera que el otro también lo sepa.

John Maynard Keynes comparó el mercado bursátil con un concurso de belleza. Tenía en mente los certámenes que eran populares en la Inglaterra de la época, cuando un periódico imprimía cien fotografías y los lectores elegían las seis caras que más les gustaban. Quienes escogían la cara que resultaba más popular participaban automáticamente en una rifa, en la cual podían ganar un premio. Keynes señaló:

> No se trata de escoger las [caras] que, según nuestro leal saber y entender, son las más bonitas, y ni siquiera las que la opinión común y corriente cree auténticamente las más bonitas. Hemos llegado al tercer nivel, donde consagramos nuestra inteligencia a prever cuál será, según la opinión común y corriente, la opinión común y corriente. Y hay algunos, me parece, que alcanzan el cuarto, quinto y sexto niveles.

En un juego inspirado en las observaciones de Keynes, se pide a los sujetos que elijan un número entre 0 y 100. El jugador cuyo número es más cercano a los dos tercios del promedio de todos los números elegidos gana un premio fijo. El promedio es obligatoriamente de 100 ó menos, por lo cual los dos tercios del promedio serán, también por fuerza, de 67 ó menos. Por ende, para cualquier promedio resultante de las elecciones de los demás jugadores, 67 estará más cerca de los dos tercios de dicho promedio que cualquier número más alto. Pero cuando los números deben ser obligatoriamente 67 o menos, los dos tercios del promedio son a la fuerza

44 o menos y así sucesivamente, hasta que se llega al equilibrio único de 0. En los experimentos, pocos sujetos escogen el 0; el número promedio ronda el 35. Quien haya elegido este número debe creer que la mayoría de los demás participantes ha elegido números más altos: el síndrome del hermano menor. El hecho de que este número sea aproximadamente los dos tercios del promedio de toda la gama, 50, sugiere que el sujeto típico cree que los otros eligen un número al azar, mientras que él tiene la libertad de optimizar. O bien el sujeto típico quizá crea que los demás pasan por dos rondas de eliminación, lo cual lo deja en libertad de optimizar mediante el agregado de una tercera ronda.

He sugerido que cuando la gente no logra ajustarse a las predicciones de la teoría de los juegos, tal vez sea porque es *menos que racional* o *más que racional*. El síndrome del hermano menor es sin lugar a dudas un fracaso de la racionalidad, como lo es la incapacidad de hacer una retroinducción simple. Ser razonable es trascender las trampas de la racionalidad: concentrarse en el hecho de que ambos jugadores pueden ganar e ignorar a la vez la lógica de la mejor respuesta. Como he dicho, esta última idea tiene alguna afinidad con la noción del punto focal, pero sólo alguna. Los puntos focales son equilibrios, en tanto que la cooperación en el dilema del prisionero repetido de manera finita, un reclamo elevado en el dilema del viajero o la elección de seguir en el juego del ciempiés no lo son. Lo que estas elecciones tienen en común con las elecciones basadas en el punto focal es una propiedad de obviedad y razonabilidad difícil de definir y muy dependiente del contexto.

Este argumento quizá parezca más similar al pensamiento mágico (Capítulo 7) que al razonamiento del punto focal. Ignorar a las sirenas de la racionalidad es seguir la conminación de John Donne en «El aniversario»:

> ¿Quiénes tan seguros como tú y yo, pues nadie puede
> Traicionarnos, excepto el uno al otro?
> Refrenemos temores, verdaderos y falsos.

Ignorar los temores *verdaderos* parece irracional o mágico. (Lo mismo vale para la ignorancia de las verdaderas perspectivas de ganancia.) De manera alternativa, y así prefiero verlo, este comportamiento refleja un criterio más alto que la mera racionalidad.[27] Estos problemas son arduos, y los lectores están invitados a decidir. El tratamiento de algunas cuestiones prosigue en el próximo capítulo.

* * *

NOTA BIBLIOGRÁFICA

Colin Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction*, Nueva York, Russell Sage, 2004, es la fuente de la mayoría de los ejemplos de este capítulo. Se encontrará un útil análisis de las condiciones en las que las predicciones de la teoría clásica de los juegos se derrumban en Jacob K. Goeree y Charles A. Holt, «Ten little treasures of game theory and ten intuitive contradictions», *American Economic Review*, 91(5), 2001, págs. 1402-1422. La idea aparentemente simple de la retroinducción resulta encerrar profundas paradojas, algunas de las cuales se exponen en la introducción de mi *The Cement of Society: A Study of Social Order*, Cambridge y Nueva York, Cambridge University Press, 1989 [*El cemento de la sociedad: las paradojas del orden social*, Barcelona, Gedisa, 1997]. El juego ilustrado en la Figura 20.2 se ha tomado de Eric Johnson y otros, «Detecting failures of backward induction: monitoring information search in sequential bargaining», *Journal of Economic Theory*, 104(1), 2002, págs. 16-47. El dilema del viajero se ha tomado de Kaushik Basu, «The traveler's dilemma: paradoxes of rationality in game theory», *American Economic Review: Papers and Proceedings*, 84(2), mayo de 1994, págs. 391-395.

Notas al pie

[21] En 2003, el presidente Chirac mostró un ejemplo de esta actitud cuando respondió a las expresiones de apoyo a la política de los Estados Unidos en Irak hechas por políticos de Europa oriental diciendo que habían perdido una gran oportunidad de callarse, y agregó que, obviamente, carecían de la educación adecuada.

[22] Digo «habitualmente» y «suelen» debido a la considerable variación de los descubrimientos. En parte, esa variación se basa en el género, y en parte, en la cultura.

[23] Una posible objeción a este análisis es que, como hemos visto en el Capítulo 15, la gente puede sentir placer al castigar a otros y, por otra parte, la expectativa del placer es tal vez lo que *motiva* el castigo.

[24] La reducción puede verse como un efecto del descuento temporal (Capítulo 6).

[25] A diferencia de lo que sucede en el juego del ultimátum, el jugador I no tiene razones para temer que el jugador II rechace la oferta de equilibrio como injustamente baja, dado que I tiene tanto la primera como la última palabra.

[26] El nombre alude a una versión del juego con cien nodos.

[27] No me refiero a la idea de *razón* (Capítulo 4), en el sentido de las motivaciones imparciales, sino a una noción que es un tanto más cotidiana. La idea (incipiente) que evoco está relacionada con el hecho de que la gente coopera con mayor frecuencia en un dilema del prisionero cuando se lo describe como un «juego comunitario» que cuando se lo caracteriza como un «juego de Wall Street».

Capítulo 21

Confianza

Bajar la guardia

El egoísmo, dijo Tocqueville, es «la herrumbre de la sociedad». De manera similar, a menudo se dice que la confianza es «el lubricante de la sociedad».[28] La vida diaria plantearía una dificultad imposible si no pudiéramos confiar en que los otros hagan lo que dicen que harán, al menos hasta cierto punto. Aunque los estudiosos han definido la confianza de diversas maneras, por mi parte utilizaré una definición simple en relación con el comportamiento: confiar en alguien es bajar la guardia, *refrenarse de tomar precauciones contra el coparticipante en una interacción*, aun cuando el otro, debido al oportunismo o la incompetencia, pueda actuar de un modo que parezca justificar las precauciones.[29] Al hablar de «oportunismo» me refiero a un egoísmo corto de vista o «crudo», no limitado por consideraciones éticas o prudenciales. Entre los actos oportunistas típicos que pueden justificar las precauciones de los otros se cuentan las mentiras, las trampas en un examen, la ruptura de una promesa, la malversación de fondos, la infidelidad conyugal o la elección de una estrategia no cooperativa en un dilema del prisionero.

Uno puede o no confiar *en sí mismo* para respetar una negociación, permanecer lejos del alcohol o mantener firme el curso de la nave cuando llaman las sirenas. La desconfianza hacia uno mismo se revela en el precompromiso o en la elaboración de reglas privadas (Capítulo 13). Sin embargo, estas estrategias pueden ser costosas a causa de los efectos de

señal. Si otros observan un ejemplo de ese comportamiento precautorio con respecto a mis *selves* futuros, tal vez infieran (incorrectamente, como hemos visto en el Capítulo 10) que carezco en general de autocontrol. Es probable, por lo tanto, que muestren renuencia a confiar en mí en las ocasiones en que 1) mi falta de autocontrol les pueda resultar costosa; 2) no haya a mano ningún dispositivo de precompromiso, y 3) las reglas privadas sean irrelevantes, como lo serían en un encuentro único. En muchas sociedades, hay normas contra la abstinencia total de alcohol, así como normas contra la ebriedad (Capítulo 22).

La *desconfianza* puede adoptar una de dos formas. Por un lado, es posible abstenerse simplemente de interactuar con un compañero potencial cuando la interacción nos hace vulnerables a la incompetencia o el oportunismo. Por otro, podemos embarcarnos en la interacción pero tomar precauciones contra esos riesgos. La confianza, en consecuencia, es el resultado de dos decisiones sucesivas: participar en la interacción y abstenerse de monitorear al copartícipe en ella. Como la decisión de no participar en la interacción es difícil de respetar, uno puede subestimar fácilmente la magnitud de la desconfianza existente en la sociedad. Podría pensarse que hay más desconfianza en una sociedad cuyos miembros se mantienen bajo observación constante unos a otros que en otra donde la gente vive más encerrada en sí misma. Sin embargo, un examen más detenido revelaría que esta última es muy ineficiente, debido a las numerosas negociaciones mutuamente beneficiosas que nunca se entablan.

Montaigne describió una respuesta confiada: «Quien tiene mi bolsa a su cuidado durante los viajes, tiénela del todo y sin control». Otros ejemplos de una muestra de confianza pueden implicar el abstenerse de actos como los siguientes:

- Leer el diario íntimo de nuestro cónyuge.
- Utilizar celadores para vigilar a los estudiantes durante los exámenes.
- Verificar las credenciales del postulante a un empleo.
- Pedir un depósito a un inquilino.
- Insistir en la firma de contratos escritos y legalmente ejecutables.
- Pedir a una pareja menos adinerada que firme un acuerdo prenupcial.
- Poner el dinero fuera del alcance de los hijos.
- Cerrar con llave la puerta de calle al salir de casa.
- Hacerse el precompromiso de castigar a los desertores en un DP.
- Pedir una segunda opinión médica o un segundo presupuesto a un mecánico de automóviles.

Como se ha señalado, el objeto de la confianza puede ser la *aptitud* de otras personas o su *motivación*. La distinción tiene una vívida ilustración en la historia de los movimientos de resistencia. En los países ocupados por los alemanes durante la Segunda Guerra Mundial, de vez en cuando se ejecutaba a miembros de la resistencia porque se los creía agentes del ocupante. También ocurría, si bien con menor frecuencia, que se los matase porque era imposible confiar en que no fueran ligeros de lengua. Una persona podía resultar ser un borracho, razón por la cual la resistencia la ejecutaba para que no revelara información peligrosa durante una borrachera. Para tomar un ejemplo más mundano, yo podría cuestionar la destreza o la honestidad de un mecánico de automóviles. Cuando pido una segunda opinión médica, a menudo se debe a que me preocupa la competencia del primer profesional, aunque también podría inquietarme que aconseje una cirugía innecesaria para llenarse los bolsillos. En lo que sigue examinaré sobre todo la cuestión de la honestidad.

Razones para confiar

Hay una serie de razones por las que la gente puede abstenerse de tomar precauciones.[30] 1) El coste de tomar precauciones podría superar los beneficios esperados, sea en una ocasión determinada o en toda la vida. Si en mi pueblo hay un mecánico de automóviles y para conseguir un segundo presupuesto tengo que viajar ochenta kilómetros en taxi, tal vez no merezca la pena hacerlo. En términos más generales, la vida siempre es demasiado corta para temer que se aprovechen de nosotros. La pérdida ocasional que resulta de confiar en quien no es digno de confianza es pequeña si se la compara con la paz espiritual que acarrea la falta de preocupaciones. 2) El acto mismo de tomar precauciones puede suministrar información susceptible de ser explotada por los oportunistas. Montaigne cita el dicho latino *«Furem signata sollicitant. Aperta effractarius praeterit»* («Las cerraduras atraen al ladrón. El infractor ignora las casas abiertas»). 3) La idea de tomar precauciones podría ser incompatible con la actitud emocional del agente hacia la otra persona. Cuando alguien está enamorado, quizá se niegue a embarcarse en el cálculo frío que implica un acuerdo prenupcial. El verso de Donne citado en el Capítulo 20 también es apropiado en este contexto: «Refrenemos temores, verdaderos y falsos». 4)

Es posible que yo tenga creencias previas sobre la confiabilidad de la otra persona. 5) Podría tratar de *inducir* confiabilidad al confiar en ella.

En lo que sigue me concentraré en 4) y más brevemente en 5). Si bien muchos estudiosos definen la confianza exclusivamente en términos de 4), creo que el énfasis en la restricción deliberada tiene la ventaja de destacar la *interacción* entre quien confía y la persona en quien se deposita la confianza. Si esta última *percibe* la falta de precauciones, la percepción podría llevarla a actuar de una manera que no habría mostrado en otras circunstancias. En el caso 2), esto sucede porque infiere que no habrá *ocasión* de exhibir un comportamiento oportunista. En otros casos, que discutiremos más adelante, la percepción puede modificar su *motivación* para comportarse en forma oportunista, como un reflejo de la intuición preanalítica de que la confianza tiene la cualidad de autocumplirse. Lo mismo es válido para la desconfianza. Como señaló Proust: «Descubiertos los celos, la persona que los inspira los considera una desconfianza que autoriza el engaño».

Razones de la confiabilidad

Hay varios fundamentos diferentes para percibir a la gente como *confiable*. Examinaré cuatro: el comportamiento pasado, los incentivos, los signos y las señales. Con frecuencia, sabemos, o creemos saber (véase el Capítulo 10), por la observación de otras personas, que éstas son coherentes en mantener sus promesas, abstenerse de mentir, tratar con cuidado los bienes ajenos, etc. Por otra parte, una persona que se sabe (poco) confiable tenderá a pensar que los demás también son (poco) confiables (el llamado efecto de falso consenso) y, por consiguiente, se inclinará a (des) confiar (de) en ellos. Como dijo La Bruyère: «Los bellacos creen fácilmente que los otros son tan malos como ellos; nada los engaña, pero tampoco pueden engañar durante mucho tiempo».[31] Hay pruebas experimentales de que, en efecto, este mecanismo funciona. A la inversa, A quizá confíe en C porque sabe que B, en quien él confía, confía en C. Sin embargo, es posible que la inferencia no sea válida, porque la confianza de B en C podría deberse simplemente al efecto de falso consenso. Como muestran estos ejemplos, a menudo confiamos o desconfiamos de la gente por malas razones, cuando creemos que los otros son más parecidos a nosotros o bien «más parecidos a

sí mismos» (más coherentes en su comportamiento) de lo que realmente son.

En la pequeña comunidad internacional de comerciantes de diamantes, donde la tentación de tener un comportamiento oportunista es enorme, un acuerdo verbal sin testigos es tan vinculante como un contrato escrito. Un comerciante que viole un acuerdo tal vez se embolse una ganancia temporaria, pero luego *todos* sus colegas lo evitarán *para siempre*.[32] Tampoco podrá legar la actividad a sus hijos, como a menudo se acostumbra en la comunidad del diamante. En el caso de los comerciantes neoyorquinos de ese rubro, la mayoría de los cuales viven en comunidades judías ultraortodoxas, un tramposo también sufriría el ostracismo social. Este último mecanismo cimenta la confiabilidad, pero no es una condición necesaria de ésta. Con frecuencia, el incentivo para mantener la reputación de honestidad y confiabilidad es suficiente.

Los signos son *rasgos* de los individuos que, acertada o erróneamente, se consideran indicativos de confiabilidad. En un estudio de los motivos que llevan a los conductores de taxis a confiar en que sus pasajeros no les robarán ni los asaltarán, se comprobó que las mujeres eran vistas como más confiables que los hombres; los mayores, que los más jóvenes; los blancos, que los negros; los más ricos, que los más pobres; los ensimismados, que los inquisitivos, y los sinceros, que los evasivos. Un taxista hispano de Nueva York consideraría más confiables a los pasajeros hispanos que a los miembros de otros grupos étnicos. Los conductores católicos de Belfast confiarían más en los pasajeros católicos que en los protestantes, y en el caso de los conductores protestantes sería a la inversa. Hay otros rasgos más genéricos, como los ojos no demasiado juntos y el hecho de mirar al interlocutor a los ojos.

Las señales son *comportamientos* que suministran pruebas de la confiabilidad. Entre ellos puede incluirse la producción deliberada o el remedo de signos. Por ejemplo, parece que una manera adecuada de generar una mirada franca es enfocarla en la base de la nariz de nuestro interlocutor. En este caso, la señal sólo funcionará si la otra persona cree que una mirada franca es un indicador confiable de la conducta e ignora lo fácil que es simularla. Otros comportamientos funcionan como señales si para los individuos poco confiables es demasiado costoso permitírselos. Para falsificar con éxito una firma quizá se requiera una prolongada práctica, mientras que poner la propia es, en esencia, un acto sin coste alguno. Un

pobre podría vestirse como un banquero de Wall Street para parecer confiable al conductor de taxi, pero es improbable que lo haga, pues los costes serían más grandes que el botín que podría esperar obtener con el robo. En contraste, la utilización del *Wall Street Journal* para llamar un taxi es algo que cualquiera puede permitirse, y por ende no discrimina entre la persona digna de confianza y la persona poco confiable. En la medida en que la confianza se apoya en la creencia de que el copartícipe de la interacción tiene un prolongado horizonte temporal (un bajo índice de descuento temporal), las exhibiciones costosas de aptitud física y esbeltez pueden actuar como una señal, dada la (falsa) creencia de que, como rasgo de carácter, la previsión existe en todos los niveles o no existe en absoluto.

A menudo confiamos en la gente porque percibimos que el egoísmo no es su única motivación. A veces, sin embargo, confiamos en las personas sólo si vemos que son egoístas. En *El halcón maltés*, el señor Gutman dice a Humphrey Bogart: «No confío en un hombre que no se ocupa de sí mismo». Napoleón afirmaba que no había que confiar en Talleyrand porque nunca pedía ningún favor para su familia. De François Mitterrand se decía que, como presidente de Francia, era igualmente desconfiado con quienes jamás le pedían favores. En términos más generales, un gran problema para los embusteros y timadores de la confianza es hacer que su víctima crea que actúan por egoísmo. Supongamos que me acerco a alguien y le digo que se puede ganar una fortuna con la inversión anticipada de una pequeña suma de dinero. Su primera pregunta será: «¿No es demasiado bueno para ser verdad?» Y la segunda será: «Si es realmente cierto, ¿por qué quiere compartir la oportunidad conmigo en vez de aprovecharla solo?» El artista consumado del timo es capaz de suscitar la confianza de su víctima contándole una historia verosímil para explicar por qué el egoísmo lo mueve a ceder parte de las ganancias. En ausencia de una historia previa de interacción, las alegaciones sobre una motivación caritativa no son creíbles.

Así como las personas pueden ser (percibidas como) más o menos dignas de confianza, también pueden ser más o menos *confiadas*. Es decir, si A y B tienen las mismas creencias sobre C (o ninguna creencia en absoluto), A tal vez confíe en C y B no. La propensión a confiar en otros es de especial importancia en la puesta en marcha de las empresas cooperativas. En interacciones repetidas, la reciprocidad puede sostener la cooperación, *pero no en el momento inicial*, cuando no hay una historia interactiva previa. Para ponerse en movimiento, las partes deben cooperar

en forma incondicional en el primer momento. Un individuo confiado adoptaría una actitud del tipo de «hoy por ti, mañana por mí»: cooperar la primera vez y reciprocar en las ocasiones ulteriores. Como dice el proverbio: «Una vez engañan al prudente y dos al inocente». La actitud de una persona desconfiada sería «hoy por mí, mañana veremos»: desertar la primera vez y reciprocar en las siguientes oportunidades. En interacciones de dos personas, se necesitan dos individuos confiados para poner en marcha la cooperación.

De cómo la confianza puede inducir confiabilidad

El hecho de que confíen en nosotros puede inducir confiabilidad, cuando la otra parte *sabe* que nos hemos abstenido de tomar unas precauciones que podríamos haber tomado. Con referencia al criado que se encargaba de su bolsa, Montaigne escribió: «Igual me engañaría aunque lo contara [el dinero]; y, si no es un demonio, oblígole a portarse bien con tan entregada confianza». Esta forma de reciprocidad difiere de la que se encarna en la estrategia del «hoy por ti, mañana por mí» en el DP. El «hoy por ti, mañana por mí» puede ser una estrategia de equilibrio cuando el juego se juega una cantidad indefinida de veces y en cualquier partida concreta las dos elecciones se hacen en forma simultánea. La observación de Montaigne, en contraste, se aplica a juegos únicos en los que una parte hace su elección y la da a conocer antes de que la otra haga la suya. Los experimentos sobre la propensión a cooperar en un DP único confirman esta intuición. En un estudio, se reveló a los sujetos la elección del otro jugador antes de que hicieran la suya. En los casos en que un jugador sabía que el otro había desertado, sólo el 3% eligió cooperar. En los casos en que sabía que el otro había escogido cooperar, cooperó el 16%.

Puede mencionarse otro experimento para destacar la textura fina de la confianza. Este juego de la confianza se juega en dos condiciones. En ambas, el «inversor» tiene la opción de transferir a un «operador» una cantidad cualquiera entre cero y diez de su dotación de diez unidades monetarias (UM). Luego, el experimentador triplica cualquier suma enviada, de modo que si el inversor transfiere diez, el operador recibe treinta. El operador puede decidir reembolsar al inversor cualquier monto desde cero hasta el total incrementado (treinta en el mismo ejemplo). Por

último, si el inversor decide hacer una transferencia, también tiene que declarar qué monto quiere que el operador le reembolse.

Estas características definen la «condición de confianza», una designación inapropiada (ésa es al menos mi opinión). En la «condición de incentivo», el inversor también tiene la opción de declarar, en el momento en que hace la transferencia y anuncia el reembolso deseado, que impondrá una multa de cuatro UM al operador si éste transfiere menos que la suma pretendida. Algunos inversores usan esta opción, otros no. Si no lo hacen, los operadores saben que el inversor contaba con esa opción pero se abstuvo de utilizarla. Según pudo verificarse en el experimento, los reembolsos más grandes se hacen en la condición de incentivo cuando no se impone multa alguna, y los más pequeños en la misma condición cuando sí hay multas; en la condición de confianza, los reembolsos se sitúan en un nivel intermedio. Este efecto fue previsto por los inversores, que invirtieron alrededor de 30% más en la «condición de incentivo, sin multas», que en cualquiera de las otras dos.

La condición de «incentivo, sin multas» corresponde a mi definición de la confianza. Preferiría dar a lo que los experimentadores llaman «condición de confianza» el nombre de *confianza ciega*. Ésta se manifiesta cuando las precauciones quedan *excluidas*, a diferencia del caso en que *no son elegidas*. Llama la atención el descubrimiento de que la confianza (no ciega) induce más cooperación que la confianza ciega. No da lo mismo bajar la guardia.

La gente puede confiar o desconfiar de las *instituciones*. Tal vez confíe en la solvencia de un banco o prefiera guardar sus ahorros debajo del colchón. Quizá confíe en la imparcialidad de un tribunal en una disputa entre vecinos o prefiera hacer justicia por mano propia. Estrictamente hablando, estos casos no se ajustan a la definición que propuse al comienzo, dado que no hay precauciones (o hay pocas) que los ciudadanos puedan tomar cuando tratan con una institución, salvo negarse a tratar con ella. Cuando confío en mi banco, manifiesto una forma de confianza ciega. En una medida marginal, mi confianza afecta la probabilidad de que el banco siga efectivamente siendo solvente, pero sólo porque mi aporte a sus reservas lo hace un tanto menos vulnerable a un pánico. En contraste, la confianza de Montaigne en su criado generaba confiabilidad por la vergüenza o la culpa que este último sentiría si la defraudaba.

Sin embargo, como argumentó el mismo Montaigne, esas consecuencias beneficiosas del hecho de confiar en otros son esencialmente subproductos (Capítulo 4). Si nos abstenemos de tomar precauciones con la finalidad exclusiva de hacer que el otro se comporte bien, es improbable que lo logremos. «Excelente medio es para ganarse el corazón y la voluntad de otros, el someterse y fiarse de ellos, con tal de que sea libremente y sin hallarse coaccionado por ninguna necesidad y a condición de que sea con confianza pura y verdadera, con la frente al menos, libre de todo escrúpulo.»

* * *

NOTA BIBLIOGRÁFICA

Edward Glaeser y otros, «Measuring trust», *The Quarterly Journal of Economics*, 115(3), agosto de 2000, págs. 811-846, ofrecen pruebas de que las personas dignas de confianza son las que confían en otros. La comunidad comercial del diamante de Nueva York es analizada por Barak D. Richman, «Community enforcement of informal contracts: Jewish diamond merchants in New York», documento de trabajo, Harvard University, Olin Center for Law and Economics, 2000, luego publicado en *Law and Social Inquiry*, 31(2), primavera de 2006, págs. 383-420. El uso de signos y señales por parte de los conductores de taxis para determinar la confiabilidad de sus pasajeros es el tema de Diego Gambetta y Heather Hamill, *Streetwise: How Taxi Drivers Establish Their Customers' Trustworthiness*, Nueva York, Russell Sage, 2005. Las maneras de hacer que una estafa parezca creíble se analizan en un libro olvidado de Arthur Allen Leff, *Swindling and Selling*, Nueva York, Free Press, 1976. El experimento sobre la cooperación en un dilema del prisionero de caso único se presenta en Eldar Shafir y Amos Tversky, «Thinking through uncertainty: nonconsequentialist reasoning and choice», *Cognitive Psychology*, 24(4), octubre de 1992, págs. 449-474. Los resultados del juego de la confianza descrito en el texto figuran en Ernst Fehr y Bettina Rockenbach, «Detrimental effects of sanctions on human altruism», *Nature*, 422, 13 de marzo de 2003, págs. 137-140.

Notas al pie

[28] En trabajos más antiguos sobre el desarrollo económico, a veces se atribuía a la *corrupción* el papel de lubricante.

[29] Así entendida, la confianza entraña una *doble abstención*: una parte se abstiene de tomar precauciones con la esperanza de que la otra se abstenga de adoptar una conducta oportunista.

[30] Uso «abstenerse» para indicar una abstención *deliberada*. En algunos de los casos que examino, la idea de tomar precauciones mediante la lectura del diario íntimo del cónyuge, por ejemplo, tal vez nunca se le haya cruzado por la cabeza a la persona en cuestión. Sin embargo, no hay aquí una «confianza ciega» tal como la definiré más adelante, si el agente ha tenido la *oportunidad* de tomar precauciones. Para el otro miembro de la relación, el hecho de que el agente tuviera la oportunidad pero no la utilizara es un signo revelador, ya se interprete la abstención como deliberada o no.

[31] La Rochefoucauld veía las cosas de otra manera: «La gente nunca es engañada con tanta facilidad como cuando se empeña en engañar a otros».

[32] No se trata de un simple mecanismo de «ojo por ojo», en el cual un jugador que deserta en una ronda y es castigado en la siguiente puede ser perdonado si vuelve a su comportamiento cooperativo. Antes bien, se trata de un mecanismo «de gatillo», por el cual una sola defección excluye la redención tras un buen comportamiento.

Capítulo 22

Normas sociales

La conciencia colectiva

Los sociólogos se refieren a veces a la «conciencia colectiva» de una comunidad, el conjunto de valores y creencias compartidos (y de los que se sabe o se cree que son compartidos) por sus miembros. En el aspecto del valor, la conciencia colectiva incluye las normas morales y sociales, la religión y las ideologías políticas. En el aspecto de la creencia, incluye las opiniones sobre asuntos fácticos así como sobre relaciones causales, que van desde los rumores acerca de la trata de blancas hasta las creencias en los efectos perversos de los beneficios del desempleo. En este capítulo considero las normas sociales y su funcionamiento. En el siguiente me ocuparé de los modos de formación colectiva o, mejor, interactiva de creencias. En mi tratamiento de los valores y las creencias hay una doble simetría. Por un lado, tengo poco que decir sobre el surgimiento de las normas sociales, no porque la cuestión sea de escaso interés, sino porque me parece demasiado ardua. Por otro lado, tengo poco que decir acerca de la sustancia de las creencias populares o colectivas. Su contenido varía enormemente en el tiempo y el espacio, mientras que los mecanismos de surgimiento, propagación, cambio y derrumbe de las creencias son más invariables.

El funcionamiento de las normas sociales

Consideremos dos enunciados:

En los días de mucho sol, siempre use ropa negra.
En los funerales, siempre use ropa negra.

El primer mandato es una cuestión de racionalidad instrumental, dado que el aire entre el cuerpo y la ropa circula con mayor rapidez cuando esta última es negra. El segundo expresa una norma social, que no tiene una significación instrumental evidente. La existencia e importancia de las normas sociales no pueden ponerse en duda. Las causas próximas implicadas en su funcionamiento son razonablemente conocidas. Sin embargo, su origen y (si la tienen) su función últimos siguen siendo controvertidos.

Una norma social es una conminación a actuar o abstenerse de actuar. Algunas normas son incondicionales: «haga X; no haga Y».[33] Entre ellas se incluyen las que imponen no comer carne humana, no tener relaciones sexuales con un hermano, no colarse en una fila, no usar nunca ropa roja (tal como algunas madres dicen a sus hijas), usar ropa negra en los funerales, comenzar con los cubiertos que están más alejados del plato o tratar en primer lugar a los pacientes más graves. Otras normas son condicionales: «si hace X, luego haga Y» o «si otros hacen X, haga X». En muchos grupos hay una norma por la cual la persona que es la primera en sugerir que se tome alguna medida queda a cargo de llevarla a cabo;[34] como consecuencia, muchas buenas sugerencias nunca se plantean. Una pareja sin hijos puede sentirse sometida a una norma conforme con la cual quien sugiera el primero tener un hijo tendrá mayor participación en su crianza; como consecuencia, es posible que algunas parejas a las que les gustaría tener hijos no los tengan.[35] Quizá no haya una norma que me indique que debo enviar tarjetas de Navidad a mis primos, pero una vez que empiezo, una norma me dice que debo continuar y otra impone a mis primos corresponder. No obstante, aunque condicionales, estas normas no lo son con respecto a ningún *resultado* que concrete la acción, como en el mandato a usar ropa negra cuando hay mucho sol.

Más adelante propondré otros ejemplos. Ante todo, sin embargo, necesito decir algo sobre la eficacia causal de las normas sociales y la diferencia que las distingue de otras normas. Una respuesta simple a la primera cuestión es que las normas sociales operan por medio de *sanciones* informales aplicadas a quienes las violan. Por lo común, dichas sanciones

afectan la situación material del transgresor, sea mediante el mecanismo del castigo directo o por la pérdida de oportunidades causada por el ostracismo social. Un granjero que viola las normas comunitarias tal vez se encuentre con su granero quemado y sus ovejas destripadas. O bien puede comprobar que su vecino rechaza su pedido de ayuda para la cosecha. El mecanismo del *chisme* puede actuar como multiplicador, al sumar sanciones de terceros al castigo original aplicado por un segundo.

Consideremos lo que podría hacer un ganadero cuando el ganado de un vecino invade repetidas veces su tierra. Tal vez se apodere del ganado, con beneficio para él y cierto coste para el vecino. Puede eliminar el ganado o reducir su valor (por ejemplo, si castra a un toro), sin beneficio para sí mismo y con algún coste para el vecino. Puede arrear la manada invasora hasta algún lugar distante, con cierto coste para él y para el vecino. O quizá corte todas las relaciones con el vecino (ostracismo). Sin embargo, esta última respuesta podría ser ineficaz, en cuanto es posible que no disuada de futuras irrupciones. La primera respuesta podría verse como una captura violenta más que como un castigo. La segunda, y sobre todo la tercera, son más adecuadas, toda vez que indican con claridad una intención de castigar, de ser necesario con cierto coste para quien aplica el castigo.

En general, sin embargo, creo que el ostracismo o la evitación son las reacciones más importantes frente a la violación de las normas. Si en vez de las invasiones reiteradas el vecino hubiera incurrido en una ruptura única de promesa, la interrupción de las relaciones habría sido la reacción más natural. Esta tesis es respaldada por la idea general de que las normas sociales funcionan a través de las *emociones* de la vergüenza en quien las transgrede y el desprecio en el observador de la trangresión (Capítulo 8). Como la tendencia a la acción correspondiente al desprecio es la evitación, que a menudo provoca pérdidas materiales a la persona rehuida, hay un vínculo entre la respuesta emocional y la imposición de sanciones. No obstante, con frecuencia estas últimas son más importantes como vehículo para la comunicación de la emoción que por sí mismas. Por otra parte, para el *sancionador*, el coste de sancionar puede ser de especial significación para comunicar la intensidad de su emoción.

La teoría de las normas sociales basada en la sanción tropieza con un problema evidente: ¿qué motiva a los sancionadores a castigar? ¿Qué encuentran en ello? De ordinario, la imposición de sanciones es costosa o riesgosa para quien sanciona. Aun cuando éste no deje de lado la

oportunidad de una interacción mutuamente provechosa, la expresión de desaprobación podría desencadenar una reacción airada y hasta violenta en quien es su objeto. Hay en este caso una importante distinción entre la desaprobación espontánea y la censura deliberada. Esta última puede ser fácilmente contraproducente y provocar más ira que vergüenza en su objeto. Aun cuando la desaprobación sea de hecho espontánea, quien es su blanco puede interpretarla, acaso de manera interesada, como una censura intencional, y reaccionar en consecuencia. Por esa razón, la imposición de sanciones es un negocio riesgoso. ¿Por qué, entonces, la gente se dedica a él?

Una respuesta podría ser que los mismos que omiten castigar se exponen al castigo. Es indudable que esto sucede. En una sociedad con fuertes normas de venganza, cabría esperar que una persona que no rehuyera a quien omite tomar venganza fuera a su vez evitada. Entre los escolares, un niño podría estar más dispuesto a interactuar con un «empollón» cuando sus compañeros no lo observan. No obstante, es improbable que un niño que se abstiene de unirse a la turba para hostigar a otro que se muestra amistoso con el empollón sufra a su vez un hostigamiento. De ahí la escasa probabilidad de que los terceros que hostigan estén motivados por el temor al castigo. En el plano experimental, la cuestión podría examinarse viendo si los terceros castigan a los copartícipes que, al aceptar ofrecimientos muy bajos en el juego del ultimátum, omiten castigar a los proponentes poco generosos. Me sorprendería que lo hicieran, y me asombraría aún más que un cuarto grupo constituido por los observadores castigara a los terceros que no castigan. A poca distancia de la transgresión original, este mecanismo deja de ser posible.

Una explicación más parsimoniosa y adecuada de la imposición de sanciones se apoya en el desencadenamiento espontáneo del desprecio y la tendencia a la acción asociada. También puede intervenir la ira, debido a la fluida distinción entre normas sociales y morales. Además, el *alarde* por la violación de las normas sociales suscitará probablemente más ira que desprecio, porque revela a otras personas que uno es indiferente a sus reacciones. Aunque los costes y riesgos de sancionar puedan mantener a raya esas tendencias a la acción espontánea, es posible que éstas sean capaces de invalidarlos. La evitación del empollón que podría ayudar a sus compañeros de clase con la tarea es costosa, como lo era en el Antiguo

Régimen la negativa de los aristócratas a permitir que sus hijas se casaran con plebeyos adinerados. Cuando un «gusto» por la discriminación adopta la forma de la negativa a emplear a mujeres o miembros de grupos minoritarios despreciados, o a comprarles, la eficiencia económica puede verse menoscabada. Con frecuencia, ese comportamiento refleja el funcionamiento de normas sociales y no de preferencias individuales idiosincrásicas, según lo muestran expresiones como «amante de los judíos» o «amante de los negros» utilizadas para condenar a quienes actúan en contra de la norma.

Lo que las normas sociales no son

Es preciso distinguir las normas sociales de una serie de fenómenos relacionados: normas morales, normas cuasimorales, normas legales y convenciones. Aunque las líneas divisorias puedan ser fluidas, hay casos bien definidos en cada categoría. Tanto las normas morales como las normas cuasimorales (Capítulo 5) son capaces de dar forma a la conducta aun cuando el agente crea que los otros no lo observan. En contraste, la vergüenza que sostiene las normas sociales es provocada por el desprecio percibido de los demás. La tendencia correspondiente a la acción es escapar de sus miradas acusadoras: esconderse, huir y hasta matarse.

Las normas legales difieren de las normas sociales en cuanto su imposición está a cargo de agentes especializados que por lo común aplican un castigo directo y no el ostracismo, pese a lo que digan los experimentos sobre el «avergonzamiento» legal. Las normas legales y sociales tienen numerosas formas de interacción. En 1990, por ejemplo, algunos legisladores estaduales de Luisiana impulsaron la reducción de las sanciones penales aplicables a quienes infligían un castigo informal a las personas que quemaban la bandera. Pese a que en 1701 un edicto autorizó a la nobleza francesa a dedicarse al comercio (sólo mayorista, no al por menor), pasaron más de cincuenta años antes de que pudieran dejarse a un lado las normas sociales que prohibían la práctica. En algunas comunidades, hay normas sociales contra la apelación a las normas legales, mientras que en otras las personas litigan a gusto y placer.

Las convenciones, o los equilibrios de convención, pueden en principio imponerse en virtud del mero egoísmo del agente, sin acción alguna de otros. Como se ha señalado en el Capítulo 19, a menudo son muy

arbitrarias. El primer día de un congreso, cada participante puede elegir su asiento más o menos al azar. El segundo día ya se ha generado una convención: todos se dirigen a los asientos que han elegido porque esa actitud es el mecanismo obvio de asignación (basado en el punto focal). El tercer día, la convención se ha consolidado como un derecho: me enojo si otro participante ha ocupado «mi» asiento. No obstante, aunque la norma social cimenta la convención arbitraria y hace más probable que ésta sea respetada, no es indispensable. Entre los neoyorquinos existe la convención de celebrar la llegada del Año Nuevo en Times Square, pero como poca gente sabrá si una persona determinada ha estado presente o no, las oportunidades de aplicar sanciones son escasas. Aun cuando las normas sociales y legales no reforzaran la norma de conducir por la derecha, los peligros en que incurriría el agente que pasara al carril izquierdo serían un vigoroso factor de disuasión.

Una categoría compleja es la de las normas legales y políticas no escritas, como las convenciones constitucionales.[36] De ordinario, éstas no tienen fuerza de ley, aunque los tribunales pueden tomarlas en cuenta en sus fallos. Son impuestas, en cambio, por las sanciones políticas, o por el temor a ellas. Hasta 1940, por ejemplo, la convención constitucional estadounidense de que nadie podía ocupar el cargo de presidente más de dos veces fue impuesta por la creencia de que quien intentara hacerlo sería derrotado. Esas normas, que son muchas, tienen alguna semejanza con las normas sociales, dado que su imposición es obra de la fuerza difusa de la opinión pública y no de organismos especializados. Hay otras convenciones políticas que se interpretan mejor como equilibrios en juegos repetidos. En muchos sistemas parlamentarios existe, por ejemplo, una convención según la cual cuando una administración termina su mandato, sus documentos internos quedan sellados y su acceso sólo se permite (a los historiadores) varias décadas después. Si bien cualquier gobierno podría sentir la tentación de abrir los archivos de sus predecesores y utilizarlos como munición política, el conocimiento de que esa actitud sentaría un precedente para que sus sucesores hicieran otro tanto basta para disuadirlo. Ésta no es una convención en el sentido expuesto en el Capítulo 19, dado que cada gobierno preferiría apartarse de ella siempre que los otros no lo hiciesen.

Normas y externalidades

Existen normas contra aquellos que imponen pequeñas externalidades negativas a muchos otros (Capítulo 17). Cuando la gente tira papeles en el parque, escupe en la calle, orina en el lago o toma café de la cafetera de la oficina sin dejar los veinticinco centavos de rigor, por lo común trata de hacerlo sin que la vean. Aunque no tema realmente las sanciones, la mera idea de que otros podrían pensar mal de ella quizá la disuada de llevar a cabo esas acciones mientras es observada. Las normas de este tipo son socialmente útiles, en el sentido fuerte de que mejoran la situación de *todo el mundo*. La norma que veda escupir en los lugares públicos es un ejemplo especialmente apropiado. Antes de que se supiera cómo se difundían las enfermedades contagiosas, el escupir era una práctica perfectamente aceptable y facilitada por la abundante presencia de escupideras. Una vez que se entendió el mecanismo del contagio, los carteles de «No escupir» aparecieron en muchos lugares públicos. En nuestros días, la norma está tan arraigada (en algunos países, al menos) que los carteles han desaparecido.

En este ejemplo, podemos observar la aparición de la norma y dar cierto crédito a la idea de que surgió *porque* era de interés público. Advertido el peligro, se creó una norma legal a la que siguió la norma social. Es más discutible que la percepción de externalidades negativas pueda crear normas sociales *sin* el paso intermedio de la intervención pública. El simple hecho de que una norma sea necesaria y se perciba que lo es, no provoca automáticamente su existencia. En los países en vías de desarrollo, no hay norma social para limitar las dimensiones de la familia. Las normas sociales contra el sobrepastoreo y la sobrepesca no surgieron de manera espontánea para impedir la tragedia de los comunes. No hay una norma que regule el uso de antibióticos, aunque su uso excesivo impone externalidades a otros a través del desarrollo de microorganismos más resistentes. Las normas contra la interpretación de música en las playas públicas y contra el uso de teléfonos celulares en las salas de concierto también deben su origen (conjeturo) a la acción de las autoridades pertinentes. Comprobamos una y otra vez que la intervención externa es necesaria para evitar que las personas se impongan unas a otras estas externalidades negativas. En algunos casos, como ocurre con la norma contra los escupitajos, la gente puede abstenerse aunque la norma desaparezca o deje de hacerse cumplir. En otros, como en la política china de «un solo hijo», parece improbable que la conducta persista si la regulación se deroga.

Grupos más pequeños quizá sean capaces de imponer estas normas sin intervención externa. En los ámbitos laborales suele haber una norma fuerte contra los trabajadores que ponen demasiado empeño en su tarea, porque se cree que sus esfuerzos podrían llevar a la dirección a disminuir la remuneración por unidad. (En este caso, la externalidad adopta la forma de un aumento de la probabilidad de un recorte de las remuneraciones.) Aunque la dirección quiera comprometerse a implementar una política de remuneraciones fijas por unidad, para inducir a los trabajadores a un mayor esfuerzo, tal vez no sea capaz de hacer una promesa creíble en ese sentido. Los rompehuelgas también son a menudo duramente sancionados por sus compañeros de trabajo. Quizá sea significativo que estos dos casos impliquen la oposición común a un adversario. En un «juego contra la naturaleza» como el pastoreo excesivo, la solidaridad no parece surgir con tanta facilidad, porque la actitud de los aprovechados no es vista como una *traición*. Desde esa óptica, no cabría esperar la aparición espontánea de una norma contra el holgazaneo en las empresas que ofrecen bonificaciones grupales, dado que su transgresión sólo perjudica a otros trabajadores sin beneficiar al «enemigo» (véase, sin embargo, el Capítulo 26).

Otras normas sociales tienen como punto de mira las externalidades negativas que un grupo de personas impone a otro. La norma contra el fumar, aun en lugares donde todavía está legalmente permitido, es un ejemplo.[37] En muchas sociedades occidentales de nuestros días, los huéspedes que fuman mucho se abstienen de hacerlo sin preguntar siquiera al anfitrión si les está permitido fumar. Lo que podríamos llamar «externalidades del ruido» subyace a la norma «a los niños debe vérselos pero no oírselos». Esta conminación podría ser en dos aspectos una norma social y no meramente una forma de castigo parental. Primero, los niños podrían rehuir a otros niños que violan la norma. Segundo, los padres podrían evitar a otros padres cuyos hijos la violan. En los compartimentos de los trenes, por ejemplo, quienes quieren imponer una «externalidad de aire fresco» a otros suelen perder la contienda a manos de quienes imponen una «externalidad de aire viciado». La razón tal vez sea que las ventanillas bajas se perciben como la opción predeterminada y, por tanto, como una base normativa.

Normas y conformismo

Las normas sociales son poco más que exhortaciones a *no asomar la cabeza*. Los habitantes de los pueblos pequeños de todas partes reconocerán la «ley de Jante», puesta por escrito en 1933 por uno que se escapó:

No creerás que *eres* algo.
No creerás que eres tan bueno como *nosotros*.
No creerás que eres más listo que *nosotros*.
No te imaginarás que eres mejor que *nosotros*.
No creerás que sabes más que *nosotros*.
No creerás que eres más importante que *nosotros*.
No creerás que *vales* algo.
No te reirás de *nosotros*.
No creerás que *interesas* a alguien.
No creerás que puedes *enseñarnos* algo.

Estas normas pueden tener muy malas consecuencias sociales. Pueden inducir a los dotados a no utilizar sus talentos, y hacer que se los califique de brujos si, a pesar de ello, los usan. También la suerte hace fruncir el ceño. Entre los bembas de Rodesia del Norte, se dice que encontrar una colmena con miel en los bosques es buena suerte; encontrar dos colmenas es muy buena suerte, y encontrar tres es brujería.

Códigos de honor

Normas fuertes y a menudo sutiles pueden regular el comportamiento en riñas, *vendettas*, duelos y, en términos más generales, la venganza. Las normas definen las acciones que exigen una represalia o una venganza, las condiciones en las cuales y los medios con los cuales éstas pueden o deben llevarse a cabo y el destino de quien no logre cumplir la norma primaria. Empezando por este último aspecto, la omisión de la venganza provoca con frecuencia una especie de muerte cívica, en la que el agente queda completamente apartado de las relaciones sociales normales. Dentro de su familia, sus opiniones no cuentan en absoluto; si se aventura fuera de su casa, enfrenta el ridículo o algo peor. Se trata de una situación paradigmática de desprecio, que genera una vergüenza intolerable.

Cualquier cosa que pueda verse, aunque sea remotamente, como un insulto para el honor del agente, puede desencadenar una represalia. En el

París prerrevolucionario, el vizconde de Ségur, un prominente calavera de mundo, se entretenía escribiendo pequeños epigramas en verso. Un rival que estaba celoso de su reputación escribió un versito en el que se burlaba sutilmente de sus composiciones. Para vengarse, Ségur sedujo a la amante de su rival y luego, cuando la mujer anunció que estaba embarazada, le dijo que sólo la había usado para desquitarse de su adversario y que ahora, habiendo alcanzado su objetivo, ya no estaba interesado en ella. (Posteriormente, la mujer murió al dar a luz.) El vizconde regresó a París y contó la historia a todo el que quisiera escucharla, sin encontrar jamás un gesto de desaprobación. Al parecer, *Les liaisons dangereuses* era una pobre imitación de la realidad.

En la Córcega decimonónica, había cuatro circunstancias que justificaban o exigían la venganza: la deshonra de una mujer, la ruptura de un compromiso, el asesinato de un pariente cercano y un falso testimonio en la justicia que provocaba la condena de un miembro de la familia. En un caso, un notario fue condenado por homicidio a raíz de un falso testimonio y al cabo de un tiempo murió en la cárcel. Su hermano se convirtió en bandolero y a lo largo de varios años mató a los catorce testigos de la acusación. En estos casos, la finalidad de la venganza es *mantener* el honor. Sin embargo, el sistema del honor también incluía acciones emprendidas con el propósito de *conquistarlo*. Montaigne se refiere a «lo que dicen los italianos cuando quieren reprochar esa osadía temeraria que se da en los jóvenes, llamándolos necesitados de honor, *bisognosi d'honore*».

En el sur de los Estados Unidos, la gente reacciona ante los insultos percibidos más vigorosamente que en el norte. Los índices de homicidios son más elevados en el sur, y la gente aprueba con mayor fuerza las reacciones de violencia ante las afrentas. En un ingenioso estudio, un cómplice del experimentador tropezaba con el sujeto, «accidentalmente adrede», y lo insultaba. Los niveles posteriores de cortisona (un reflejo de las reacciones al incidente) y testosterona (un reflejo de la preparación para una futura agresión) exhibieron un crecimiento espectacularmente más grande en los sujetos sureños que en los sujetos norteños. En otro experimento, los sujetos seguían caminando por el corredor donde habían tenido el «choque» y veían avanzar hacia ellos, de manera resuelta, a un tipo con aspecto de jugador de fútbol americano (uno preparado para actuar). Como el corredor estaba obstruido con mesas, sólo podía pasar una persona por vez, lo que generaba en esencia un juego de la gallina. Antes de

«achicarse», los sureños se acercaban mucho más a la otra persona (un metro) que los norteños (tres metros).

¿Los códigos de honor cumplen alguna función social? Si la cumplen, ¿puede la función explicar su existencia? La idea de que la práctica de la venganza es una forma útil de control de la población es demasiado arbitraria para tomarla en cuenta. Un punto de vista alternativo, a saber, que las normas de la venganza proporcionan un equivalente funcional de la imposición organizada de la ley en las sociedades con un Estado débil, también es poco verosímil. Las sociedades mediterráneas y de Medio Oriente que han adherido a esas normas han tenido niveles de violencia e índices de mortalidad entre hombres jóvenes muy superiores a los que se constatan en otros lugares.[38] Según lo sugiere la observación de Montaigne citada hace un momento, las normas de la venganza y el código de honor más amplio en el cual están inmersas pueden *encender tantos fuegos como los que apagan*. Con frecuencia, las riñas provocan más disrupción que control.

Otros han sostenido que las normas del honor se desarrollan en sociedades pastorales de población dispersa, en las cuales la fama de apelar de buena gana a la violencia funciona como un factor disuasivo útil y hasta indispensable del robo. La cultura del honor en el sur norteamericano se ha explicado desde esa perspectiva. Más allá de los problemas generales de la explicación funcional, este análisis tropieza con la dificultad de que los códigos de honor eran igualmente fuertes en la corte de los reyes franceses de los siglos XVII y XVIII, para mencionar un solo ejemplo no rural. Algunos de los que hacen hincapié en los códigos de honor en la aristocracia urbana y no entre los pastores rurales dan entonces con otra explicación funcional: a falta de guerra, la nobleza «necesitaba» duelos para mantener su espíritu belicoso. Si no se hace alusión a un mecanismo por medio del cual la necesidad genere su propia satisfacción, ese argumento no vale nada. Estos comentarios polémicos no implican que yo pueda proponer una explicación más adecuada.

Normas formales de educación

Otro conjunto de normas sociales es el incluido en las reglas sobre los modales o *normas formales*. Los códigos de vestimenta, lenguaje, maneras de la mesa y cosas por el estilo son a menudo implacables en el detalle, y

condenan al ostracismo a quienes pasan por alto el más mínimo de los matices.[39] En todas las sociedades hay una norma que regula la distancia apropiada que es preciso mantener con otras personas en las ocasiones sociales. Si uno penetra en el espacio privado de una persona (en los Estados Unidos, tal vez unos cuarenta centímetros), corre el riesgo de ser rehuido por su grosería. La norma, sin embargo, tiene la característica común de que los individuos en cuestión ignoran con frecuencia su existencia y su funcionamiento. La mayoría de las normas formales están sumamente codificadas, a menudo al pie de la letra. No sólo son inútiles (en su mayor parte), sino también, en ocasiones, crueles en sus consecuencias, como sucede cuando una niña de cinco años vuelve a su casa bañada en lágrimas porque sus amigas han ridiculizado su nuevo cochecito para la muñeca con el argumento de que *no tiene frenos*. En el París prerrevolucionario, un joven oficial, adinerado pero no noble, trató de colarse en un baile en Versalles. «Lo trataron con tanta severidad que, en su desesperación por el ridículo con el que se había cubierto, en una época en que el ridículo era el peor de todos los males, se mató en el camino de regreso a París.»

El enigma es por qué estos asuntos intrínsecamente triviales asumen tamaña importancia. La desaprobación desmedida suscitada por una ruptura de la norma puede deberse a la infundada creencia de que la gente está hecha toda de una pieza (Capítulo 10), de modo que alguien que transgrede una norma poco importante también se inclinará a violar otras de mayor trascendencia. Además, la violación de las triviales normas de forma puede verse como una muestra no trivial de desdén por lo que piensan otras personas. Sin embargo, con ello no se explica por qué existen las normas poco importantes. El enigma no es por qué existe tal o cual norma, sino por qué la gente habría de atribuir alguna significación a asuntos intrínsecamente intrascendentes. No obstante, una vez que se las *ve* como importantes, las normas *se convierten* en importantes reguladores del comportamiento. Si alguien que tiene una entrevista laboral en un banco se presenta vestido con una chaqueta de cuero de color rosa, lo único que puede deducirse es que o bien se mofa deliberadamente de la norma social o carece de sensibilidad a las expectativas de los otros. En uno y otro caso, su comportamiento brinda una buena razón para dar el empleo a otra persona.

Insistamos en que las explicaciones funcionales son muy comunes. Las sutiles reglas formales de educación en la elite existen, presuntamente, a fin

de dificultar a las personas ajenas a ella «colarse en la fiesta» gracias a la imitación del comportamiento regido por aquéllas. Es indudable que, con frecuencia, dichas reglas tienen el *efecto* de reprimir a los advenedizos, pero con ello no explicamos por qué existen. Como lo han verificado muchos estudiantes proletarizados por decisión propia, para quien no ha nacido en la clase obrera es muy difícil entrar a ella. En Noruega, por ejemplo, los jóvenes maoístas comprobaron durante la década de 1970 que las burlas dirigidas a la familia real eran una manera segura de granjearse la desafección de la clase a la que procuraban incorporarse. Sin embargo, a nadie se le ha ocurrido sugerir que las normas de la clase obrera existen *a fin de* dificultar a las personas ajenas a ella hacerse pasar por trabajadores. El argumento no tiene más sentido para las normas de la elite.

Normas que regulan el uso del dinero

Además de las proscripciones legales al uso de dinero para la compra de bebés, votos, calificaciones escolares, órganos para trasplantes y (a veces) sexo, hay muchas normas sociales que regulan sus usos apropiados e inapropiados. Algunas de ellas tienen vigencia entre amigos y vecinos, como la que impide pedir a otro adulto de una comunidad suburbana que corte nuestro césped por dinero (Capítulo 12). Cuando los vecinos tienen que colaborar en la construcción de un cerco, a menudo hacen contribuciones en especie: es posible que uno suministre la mano de obra y otro las materias primas. Aun cuando quizá sea más eficiente que uno de ellos haga todo el trabajo y el otro le reembolse su parte, la norma contra las transacciones económicas entre vecinos excluye esta solución. En los Estados Unidos, los profesores de las universidades privadas ni siquiera soñarían, a menos que fueran amigos íntimos, en preguntarse unos a otros cuánto ganan.

Acaso de manera sorprendente, otras normas regulan el uso del dinero en las relaciones entre extraños. Hay una norma, por ejemplo, que proscribe abordar a la persona que está a la cabeza de la fila en una parada de autobús y ofrecerle dinero para tomar su lugar.[40] Esta norma es obviamente ineficiente: si la persona solicitada acepta ir al final de la fila a cambio de dinero, ambos agentes se benefician y nadie se perjudica. Según Tocqueville, esas normas contra la ostentación de riqueza en público son específicas de las sociedades democráticas: «¿Veis a este opulento

ciudadano? […] Su apariencia es sencilla y su andar modesto. Entre las cuatro paredes de su casa se adora el lujo». Hay normas que incluso regulan la conversación sobre el dinero o la conciencia del aspecto monetario de una transacción. Los menús dobles que aún se presentan a las parejas en algunos restaurantes, uno sin precios para la mujer y otro con ellos para el hombre, refleja una norma según la cual la galantería no debe mancharse con cuestiones monetarias. Cuando llevamos una botella de vino a una fiesta, se supone que le sacamos la etiqueta con el precio.

Normas del beber

Si las normas sociales se encaminaran invariablemente a mejorar el bienestar del individuo o la sociedad, cabría esperar que proscribieran los excesos en la bebida, considerados como de consecuencias nocivas a corto o largo plazo. Hay, en efecto, muchas normas de este tipo. Algunas, por lo común vinculadas a la religión, exigen total abstinencia. El islam y algunas sectas protestantes tienen prohibiciones absolutas con respecto al alcohol. Las normas seculares, en contraste, a menudo prescriben beber con moderación. La norma italiana «nunca bebas entre comidas» tiene el efecto dual de limitar el consumo total y reducir la velocidad de absorción del alcohol, con lo cual amortigua el efecto de corto plazo sobre el cuerpo. En Islandia, hay normas que vedan beber en presencia de los niños o durante las excursiones de pesca.

Sin embargo, las normas relacionadas con el alcohol no siempre mejoran el bienestar. Hay normas que condenan la abstinencia, así como las hay que prescriben a la gente beber mucho. Entre los indios mapuches de Chile, se critica a quienes beben solos, al igual que la abstinencia; se estima que un comportamiento semejante es una muestra de falta de confianza. La cultura francesa tradicional condena tanto al sobrio como al borracho. En Italia, la desconfianza que rodea a los abstemios se expresa en el proverbio «protéjame Dios de quienes no beben». En las subculturas juveniles de muchos países, quienes se abstienen se ven sometidos a una fuerte presión y el ridículo. A la inversa, hay numerosas sociedades donde la abundancia en el beber está socialmente prescrita. En México y Nigeria, las cualidades machistas demostradas en la capacidad de beber mucho son muy admiradas. En la Rusia prerrevolucionaria, la ingesta excesiva de alcohol era obligatoria en la subcultura de los oficiales jóvenes.

Cuando se condena la abstinencia o la bebida abundante es socialmente obligatoria, es posible que los aspirantes a abstinentes tengan que recurrir a subterfugios. En Suecia, una pregunta habitual es: «¿Quieres jerez, o vas a conducir?» Se acepta de tal modo que los alcohólicos abstinentes respondan a menudo que van a conducir, porque esto los libera de la presión social que, de lo contrario, el anfitrión seguramente ejercería para convencer al huésped de tomar un trago. La norma de la bebida sólo puede ser contrarrestada por otra norma (que prohíbe conducir borracho). De manera similar, se ha sostenido que las conversiones al protestantismo brindan una alternativa a algunos latinoamericanos que quieren apartarse del sistema de gobernanza comunitaria en el cual aun los rituales suelen implicar una abundante ingesta alcohólica y borracheras. Una vez más, la norma de la bebida sólo puede ser invalidada por otra norma, que en este caso cuenta con el respaldo de la religión.

Estos son casos de uso estratégico de normas. A la inversa, la gente puede comportarse estratégicamente para eludirlas. Algunos antiguos chinos consideraban que el alcohol mismo era sagrado y sólo lo bebían en ceremonias sacrificiales; a la larga, terminaban por hacer sacrificios cada vez que querían beber. En España, a determinadas horas, no beber con el estómago vacío es una proscripción cultural tácita, de modo que junto con la bebida hay que incluir comida. En ambos casos observamos una inversión del vínculo causal original: en vez de obedecer la norma condicional de beber únicamente cuando hace X, la gente hace X cada vez que quiere tomar un trago.

La norma de la propina

La propina que se deja por un servicio no es un fenómeno desdeñable. Los cálculos de las propinas dejadas en los restaurantes estadounidenses oscilan entre cinco mil y veintisiete mil millones de dólares por año; si se agregaran las propinas a taxistas, peluqueros y otros, se llegaría a una cifra más grande. Las estimaciones de la proporción de los ingresos que los camareros obtiene con las propinas van del 8% (el supuesto del Servicio de Ingresos Internos [IRS]) hasta el 58% en el caso de quienes sirven un menú completo. En algunos contextos, el hecho de dejar propina puede ser enigmático; en otros, no tanto. Si nos atendemos con el mismo peluquero cada vez que necesitamos un corte de pelo, la propina nos asegura un buen

servicio; lo mismo vale para las comidas en nuestro restaurante favorito. La propina dejada en encuentros irrepetibles, como un viaje en taxi o la comida en un restaurante al que no esperamos volver, es más enigmática. De hecho, esa conducta es doblemente enigmática: no puede ser sostenida por interacciones bipartitas a lo largo del tiempo ni por sanciones de un tercero en el momento del encuentro. Si somos el único pasajero en el taxi, otras personas difícilmente estarán en condiciones de saber si le hemos dado una propina adecuada al conductor; tampoco es probable que los demás clientes del restaurante adviertan cuánta propina le dejamos al camarero.

Se ha sostenido que las propinas son una manera eficiente de remunerar a los camareros. La supervisión de la calidad del servicio es una tarea evidentemente más fácil para el cliente que para el dueño del restaurante. Por ello, la descentralización de esa función y la conexión de la recompensa con el desempeño observado son una manera de superar el «problema del director y los agentes» (cómo impedir que los trabajadores holgazaneen) que afecta muchas relaciones contractuales (Capítulo 26). Las propinas, por lo tanto, podrían formar parte de un «contrato implícito» con la finalidad de mejorar la eficiencia. Pero como dijo Sam Goldwyn, un contrato no escrito no vale el papel en el que está escrito. El argumento, como muchos otros intentos de explicar las normas sociales, no es más que un ejemplo de funcionalismo sin respaldo. La idea de que los propietarios de restaurantes que prohíben las propinas quedan eliminados en la competencia con aquellos que las permiten es completamente conjetural, y en todo caso no explicaría por qué los clientes sí dejan propina en estos últimos. Además, cuando se somete a una evaluación empírica, el hábito de las propinas no parece pasar las pruebas pertinentes de eficiencia. No parece, por ejemplo, ser más preponderante en las ocupaciones en que la supervisión es más fácil. El hecho de que los camareros a menudo hagan un fondo común con sus propinas también debilita el argumento de la eficiencia.

No sé por qué hay una norma para dejar propina en ciertas ocupaciones y no en otras. Sin embargo, una vez que una norma existe, podemos entender por qué la gente deja propinas: sencillamente no le gusta la idea de que otros, como un taxista decepcionado, puedan desaprobarla, aun cuando no espere volver a encontrarse con ellos. No es necesario ser el objeto de una mirada despreciativa del otro. Tal vez baste simplemente con saber o tener razones para creer que el otro siente desprecio. Para tomar otro ejemplo, la creencia de que los otros podrían desaprobarme explica por qué

me abstengo de meterme el dedo en la nariz en el andén del metro cuando un tren pasa sin detenerse, aunque no haya otras personas en ninguno de los dos andenes.

¿Por qué las normas?

Como ya he dicho, la importancia de las normas sociales para la regulación del comportamiento y el mecanismo próximo mediante el cual funcionan se entienden bastante bien. No creo, no obstante, que tengamos una buena comprensión de sus orígenes. Hay al respecto dos interrogantes independientes. Primero, ¿cuál es el origen evolutivo de las emociones correlativas de la vergüenza y el desprecio que sostienen las normas sociales? En otras palabras, ¿por qué hay sencillamente normas sociales? Segundo, ¿por qué existen normas específicas en sociedades específicas? ¿Cómo y cuándo aparecen? ¿Cómo y cuándo desaparecen?

Una respuesta simple a la primera pregunta es que nos preocupamos intensamente por lo que otras personas piensan de nosotros. Buscamos su aprobación y tememos su desaprobación. Sin embargo, esta respuesta no hace más que plantear la misma pregunta a un paso de distancia: ¿por qué debemos preocuparnos por lo que la gente piensa de nosotros? En algunos casos, claro está, una reputación puede ser útil y digna de cultivarse. No obstante, la idea de que el conductor del taxi podría pensar mal de nosotros si no le dejamos una propina está completamente divorciada de las inquietudes de la reputación. Por otra parte, como la razón por la que otros piensan mal de nosotros es que hemos violado una norma social, la explicación de las normas por el deseo de que los otros no piensen mal de nosotros es hasta cierto punto circular.

En lo concerniente a la segunda pregunta, la respuesta más habitual es que las normas surgen para regular las externalidades. La idea tiene algo de cierto si agregamos, como he sostenido que debemos hacerlo, que las normas sociales contra la imposición de externalidades negativas a otros suelen ser introducidas por una autoridad externa. Hay una norma social general por la cual debemos obedecer la ley. Si las multas se concibieran como precios, y la prisión no fuera más estigmatizante que una estadía en el hospital, esa norma no existiría, pero en general estas reacciones ante la transgresión de la ley no se ven como equivalentes a otras obligaciones objetivamente iguales. La gente se avergüenza de ir a la cárcel y trata de

ocultarlo si puede.[41] Cuando la ley proscribe conductas que imponen externalidades negativas a otros, la norma social de obedecerla puede derivar en otra contraria a dicha conducta. La norma tal vez persista aun cuando la ley que le ha dado origen caiga en desuso. Con todo, este resultado puede ser difícil de distinguir del surgimiento del «buen equilibrio» en un juego de la seguridad (Capítulo 19). Si para inducir cooperación el Estado castiga a los desertores, pero luego desmantela el aparato de castigo, es posible que la gente siga cooperando porque la situación más jerarquizada por cada uno es aquella en la cual él mismo y todos los demás cooperan (no existe la tentación de ser un aprovechado).

Con respecto a muchas de las demás normas que he examinado, como la que veda ofrecer dinero para comprar el lugar de alguien en la cola del autobús, las normas formales de educación o las normas sobre la propina, cuesta más dar con una explicación de su aparición y persistencia. Una línea argumental, a menudo propuesta por los economistas, es que la persistencia de las normas puede explicarse como un comportamiento de equilibrio, y que su aparición es una cuestión accidental e histórica acerca de la cual la ciencia social tiene poco que decir. Habida cuenta de que una premisa de este libro es que la línea divisoria entre las ciencias sociales y la historia es artificial y carece de sentido, no puedo coincidir con esa última idea. En cuanto a la primera, he sostenido que el rasgo típico de las normas sociales es no mostrar la lógica de la mejor respuesta que caracteriza los juegos estratégicos. Cuando, sin que me observen, veo a otro violar una norma, la sanción del transgresor no es habitualmente la mejor respuesta.

* * *

Nota bibliográfica

Este capítulo se apoya en y mejora (espero) la descripción de las normas que he propuesto en *The Cement of Society: A Study of Social Order*, Cambridge y Nueva York, Cambridge University Press, 1989 [*El cemento de la sociedad: las paradojas del orden social*, Barcelona, Gedisa, 1997], y, de manera más sucinta, en «Social norms and economic theory», *Journal of Economic Perspectives*, 3(4), 1989, págs. 99-117. Entre los análisis más influyentes de las normas sociales deben mencionarse James Samuel Coleman, *Foundations of Social Theory*, Cambridge (Massachusetts), Harvard University Press, 1990; Robert C. Ellickson, *Order without Law:*

How Neighbors Settle Disputes, Cambridge (Massachusetts), Harvard University Press, 1991, y Eric A. Posner, *Law and Social Norms*, Cambridge (Massachusetts), Harvard University Press, 2000. Aprendí de todos ellos, pero ninguno me convenció. Se encontrará una instructiva crítica de Posner en la reseña de Richard H. McAdams, «Signaling discount rates: law, norms, and economic methodology», *Yale Law Journal*, 110(4), enero de 2001, págs. 625-690. Hay provechosas discusiones de las normas o convenciones constitucionales no escritas en dos artículos de Joseph Jaconelli, «The nature of constitutional convention», *Legal Studies*, 24(19), 1999, págs. 24-46, y «Do constitutional conventions bind?», *The Cambridge Law Journal*, 64(1), 2005, págs. 149-176. Tomo la ley de Jante de Aksel Sandemose, *A Fugitive Crosses His Tracks*, Nueva York, Alfred Knopf, 1936. El papel de la brujería en el mantenimiento de las normas que vedan asomar la cabeza se examina en Keith V. Thomas, *Religion and the Decline of Magic: Studies in Popular Beliefs in Sixteenth-and Seventeenth-Century England*, Harmondsworth, Penguin, 1973. Analizo los códigos de honor y venganza en el Capítulo 3 de *Alchemies of the Mind: Rationality and the Emotions*, Cambridge y Nueva York, Cambridge University Press, 1999 [*Alquimias de la mente: la raciona lidad y las emociones*, Barcelona, Paidós, 2002]. La historia sobre el vizconde de Ségur se ha extraído de Éléonore-Adèle d'Osmond, condesa de Boigne, *Les Mémoires de la comtesse de Boigne, née Osmond: récits d'une tante*, vol. 1, París, Mercure de France, 1999, págs. 73-74. Estas memorias (*ibíd*., pág. 38) son también la fuente de la historia del joven oficial que se suicidó por la vergüenza de caer en el ridículo. Las normas formales de educación son el tema de Pierre Bourdieu, *Distinction: A Social Critique of the Judgment of Taste*, Cambridge (Massachusetts), Harvard University Press, 1987 [*La distinción: criterios y bases sociales del gusto*, Madrid, Taurus, 1991], que muestra un sesgo característicamente funcionalista. Los estudios experimentales sobre la «cultura del honor» se presentan en Richard E. Nisbett y Dov Cohen, *Culture of Honor: The Psychology of Violence in the South*, Boulder (Colorado), Westview Press, 1996. Los ejemplos de las normas sobre la bebida son de mi *Strong Feelings: Emotion, Addiction, and Human Behavior*, Cambridge (Massachusetts), The MIT Press, 1999 [*Sobre las pasiones: emoción, adicción y conducta humana*, Barcelona, Paidós, 2001]. Las desventuras de los estudiantes proletarizados por decisión propia en Noruega se exploran en una novela maravillosamente entretenida, por

desdicha no traducida al inglés, de Dag Solstad, *Gymnaslærer Pedersens beretning om den store politiske vekkelsen som har hjemsøkt vårt land*, Oslo, Oktober, 1982. Sobre las normas que prohíben preguntar a otra persona cuánto gana, véase Matthew Edwards, «The law and social norms of pay secrecy», *Berkeley Journal of Employment and Labor Law*, 26(1), 2005, págs. 41-63. La explicación basada en la eficiencia de la norma sobre las propinas es propuesta por Nancy L. Jacob y Alfred N. Page, «Production, information costs and economic organization: the buyer monitoring case», *American Economic Review*, 70(3), 1980, págs. 476-478. Se la critica en Michael Conlin, Michael Lynn y Ted O'Donoghue, «The norm of restaurant tipping», *Journal of Economic Behavior & Organization*, 52(3), 2003, págs. 297-321, que propone una explicación más cercana a la esbozada aquí.

Notas al pie

[33] En lo que sigue, «condicional» e «incondicional» se refieren al contenido de la norma social. Como se ha señalado en el Capítulo 5, todas las normas sociales son condicionales, en cuanto su funcionamiento es contingente de la presencia de un observador.

[34] Esta norma puede vincularse al razonamiento del punto focal.

[35] La norma, por lo tanto, induce un juego de la gallina.

[36] La expresión «convención constitucional» se utiliza con referencia a dos aspectos enteramente diferentes de las constituciones: las normas no escritas que complementan la constitución escrita y las asambleas constituyentes que suelen adoptarse como mecanismo para sancionar una de éstas.

[37] La externalidad más importante es generada por la inhalación de humo (fumar pasivo). A veces también se sostiene que los fumadores imponen una externalidad negativa a otros fumadores que quieren dejar el hábito, pero no pueden resistir el deseo de fumar suscitado por la indicación visual de otras personas que fuman.

[38] Podría objetarse que la comparación pertinente es con el nivel de violencia que existiría en el «estado de naturaleza». Si este estado se define por un interés exclusivo en el egoísmo y la ausencia de todo órgano estatal, no generaría violencia motivada por la envidia, el despecho o la ira. La previsión de que los fuertes han de apropiarse enérgicamente de los bienes producidos por los débiles podría tener un efecto paralizante sobre la producción e impedir así la aparición de la violencia concreta.

[39] En las sociedades aristocráticas, a veces se aceptan las desviaciones *groseras*, vistas como deliberadas y no como una muestra de la ignorancia de las reglas. El Charlus de Proust es un ejemplo.

[40] Si la persona dijera que necesita llegar rápidamente a su casa porque tiene un hijo enfermo, tal vez se le permitirá ponerse a la cabeza de la fila, pero aquel cuyo puesto ocupaba quedará habitualmente en segundo lugar. Como este acto impone costes a todas las otras personas que esperan el autobús, también es preciso obtener de algún modo su aprobación.

[41] En Noruega se imponía una sentencia obligatoria de tres semanas de cárcel a quienes condujeran ebrios. Algunas personas llevaban a la prisión una lámpara de rayos ultravioletas para adquirir un bronceado que pudiera dar crédito a su historia de que habían estado de vacaciones.

Capítulo 23

Formación colectiva de creencias

Tocqueville sobre el conformismo

Los mecanismos de formación de creencias que consideré en el Capítulo 7 funcionan principalmente en el nivel del individuo, en el sentido de que las creencias sostenidas por una persona deben poco a las sostenidas o expresadas por otras. En este capítulo examino algunos mecanismos de formación colectiva o interactiva de creencias. Para ilustrar la distinción, veamos los análisis de Tocqueville sobre el conformismo estadounidense. Una explicación de la causa por la cual los estadounidenses suelen tener las mismas ideas es simplemente que viven en condiciones similares: como hombres «de condición semejante […], miran los objetos bajo el mismo aspecto, su espíritu se inclina naturalmente hacia las mismas ideas, y aunque cada uno pudiera separarse de sus contemporáneos y formar creencias particulares, acaban por encontrarse todos, sin saberlo y sin querer, en cierto número de opiniones comunes». Otra explicación se apoya en la presión para conformarse: «En Norteamérica, la mayoría traza un círculo formidable en torno al pensamiento. Dentro de esos límites el escritor es libre, pero ¡ay si se atreve a salir de él! No es que tenga que temer un auto de fe, pero está amagado de sinsabores de toda clase, de persecuciones todos los días».

Este último pasaje sugiere que la gente se conforma en lo exterior, debido a la presión social, pero no necesariamente en su fuero íntimo. Según escribe también Tocqueville, si se defiende una opinión desviada,

«cuando os acerquéis a vuestros semejantes, huirán de vosotros como de un ser impuro; y quienes creen en vuestra inocencia, esos mismos os abandonarán, porque huirán de ellos a su vez». Otros pasajes indican que el conformismo llega hasta el alma, de modo que, a la larga, la gente se forjará una creencia sincera en la opinión mayoritaria. Se sugieren dos mecanismos, uno «frío» o cognitivo y otro «caliente» o motivacional. Por un lado, «no pueden concebir que, teniendo todos luces iguales, no se encuentre la verdad al lado del mayor número». Por otro, el hecho de que «entre los norteamericanos las leyes políticas [sean] tales que la mayoría rige soberanamente la sociedad […], aumenta demasiado el imperio que ejerce sobre la inteligencia, porque nada hay más común en el hombre que reconocer una ciencia superior en aquel que lo oprime».

Descubrimientos experimentales

He citado a Tocqueville con cierta extensión (y volveré a citarlo en este capítulo) debido a su agudo discernimiento de estos asuntos. Las cuestiones que él identificó –conformismo externo contra conformismo interno, y mecanismos cognitivos contra mecanismos motivacionales– nos siguen acompañando en nuestros días. Para abordarlas, me referiré en primer lugar a algunos experimentos clásicos sobre la conformidad.

En el más célebre de esos experimentos, se pidió a los sujetos que indicaran cuál de tres líneas, A, B y C, se aproximaba más en longitud a una cuarta, D. Había tres condiciones: privada, doblemente pública e individualmente pública. En la condición privada, los sujetos daban su respuesta cuando no había nadie presente al margen del experimentador. En este caso, el 99% indicó que B era la más semejante a D, lo cual sugiere la corrección inequívoca de esta respuesta. Sin embargo, en las dos condiciones públicas una minoría sustancial de los sujetos dio diferentes contestaciones. En ambas condiciones, el sujeto respondía luego de que varios otros (cómplices del experimentador) hubieran dicho en forma unánime que A era la de longitud más aproximada. En la condición doblemente pública, en la que el sujeto daba su respuesta en presencia de los cómplices, alrededor de un tercio coincidió en que A era la más semejante.[42] En la condición individualmente pública, en la cual los sujetos expresaban su opinión en privado luego de haber oído lo que decían los otros, el conformismo se redujo sin eliminarse.

El excesivo conformismo exhibido en la condición dos veces pública se debía posiblemente al *temor a la desaprobación*. El conformismo residual en la condición individualmente pública podía deberse al *aprendizaje* («es improbable que tantos estén equivocados») o a la *reducción de la disonancia*. Esta última explicación parece más verosímil. Es improbable que quienes manifestaban en privado su coincidencia con la mayoría lo hicieran sobre la base exclusiva del aprendizaje racional, dado el débil estatus cognitivo de la opinión mayoritaria. Debía de intervenir algún factor motivacional.

Otro experimento refuerza esta interpretación. En este caso, los sujetos tenían una tarea más ambigua: detectar la distancia que había recorrido una fuente de luz en una habitación oscura. Aunque la fuente estaba en realidad inmóvil, los sujetos aislados juzgaron que había recorrido unos diez centímetros («efecto autocinético»). Tras oír a un cómplice del experimento decir que la luz se había movido entre treinta y ocho y cuarenta centímetros, los sujetos calcularon que la distancia era de aproximadamente veinte centímetros. Cuando dos de los simuladores hicieron estimaciones en torno de los cuarenta centímetros, el cálculo de los sujetos fue de unos treinta y cinco centímetros. Es decir que la presencia de un cómplice del experimento provocaba un aumento de diez centímetros en el cálculo, y la de un segundo asociado del experimentador, un incremento adicional de otros quince centímetros.

En un proceso de aprendizaje bayesiano (Capítulo 11), puedo apoyarme en otros observadores para corregir mi percepción o mi recuerdo. Sus estimaciones de algún hecho, como la distancia recorrida por la luz, pueden servir para modificar mi evaluación inicial. *Cuánto* la afectarán dependerá de mis creencias sobre la confiabilidad de su percepción y de la cantidad de dichos observadores. En este experimento, el sujeto atribuiría presumiblemente la misma confiabilidad a cada cómplice. Cualquiera fuera esa confiabilidad, el cambio en su cálculo originado en la afirmación de un cómplice de que la distancia era de cuarenta centímetros debía ser más grande que el cambio adicional causado por el segundo cómplice.[43] Sin embargo, esta situación contradice los descubrimientos, pues el segundo cómplice causó un *mayor* ajuste que el primero. Parece haber un efecto de reducción de la disonancia, provocado por la incomodidad de verse en discrepancia con la mayoría, que no puede limitarse al aprendizaje racional.

El segundo experimento tuvo una nueva e interesante característica. Se extendió a lo largo de varias «generaciones», en cuyo transcurso las personas preparadas para el experimento fueron gradualmente reemplazadas por sujetos ingenuos. Así, en la segunda generación de un experimento con dos cómplices, el lugar de uno de ellos fue ocupado por un sujeto ingenuo de un experimento de la primera generación, mientras que en la tercera también se reemplazó al otro cómplice por un sujeto ingenuo de una generación anterior. En generaciones ulteriores, todos los participantes eran sujetos ingenuos que con anterioridad habían estado en contacto con personas preparadas para el experimento o con otros sujetos que lo habían estado, y así sucesivamente. El experimento se diseñó de manera tal que el nuevo sujeto reclutado en cada generación hablara después de los otros dos. Los diseñadores habían previsto que los cálculos artificialmente altos se mantendrían de manera indefinida, pero estaban equivocados. Al cabo de unas seis generaciones en grupos de tres personas y de ocho generaciones en grupos de cuatro personas, las estimaciones convergieron en los diez centímetros, esto es, el cálculo de la distancia dado por los sujetos aislados. La creencia en el nuevo traje del emperador no se perpetuó indefinidamente. El mantenimiento a lo largo del tiempo de algunas creencias culturales con escaso respaldo en la realidad podría deberse a que la discrepancia es difícil de observar o a que se apoyan en otros fundamentos. El uso de loterías para identificar buenos sitios de caza o pesca, como es de práctica en algunas sociedades, quizás haya sobrevivido a raíz de su significación religiosa.

Ignorancia pluralista

Al comienzo de este capítulo he distinguido entre dos razones por las cuales las personas podrían en un momento dado sostener o profesar creencias similares: porque se encuentran bajo la influencia de condiciones similares (correlación) o porque influyen unas sobre otras (causación). Un caso especial de la primera situación es el que muestran los abundantes ejemplos de descubrimientos simultáneos, como la invención del cálculo por Newton y Leibniz más o menos en la misma época. Aunque nadie sabe exactamente cuáles eran las «condiciones similares» en ese caso, es posible que la idea estuviera «en el aire». Otro caso de aparición simultánea de ideas similares es el del nuevo traje del emperador. El cuento de Hans

Christian Andersen se publicó en 1835. En el segundo volumen de *La democracia en América*, publicado en 1840, Tocqueville dio con una idea parecida para explicar la estabilidad aparente de la opinión mayoritaria:

> Algunas veces sucede que el tiempo, los acontecimientos o el esfuerzo individual o aislado de las inteligencias, acaban por conmover o destruir poco a poco una creencia, sin que se descubra nada en lo exterior. No se la combate ciertamente, ni se reúne nadie para hacerle la guerra. Sus sectarios empiezan a dejarla uno a uno sin ruido; pero cada día la abandonan algunos, hasta que al fin no la sigue más que un corto número, y en ese estado reina todavía. Como sus enemigos continúan en silencio, o si se comunican es en secreto, se hallan por mucho tiempo sin saber que se efectúa una revolución, y en esta duda permanecen inmóviles, observan y callan. La mayoría no cree, pero finge creer, y ese vano fantasma de la opinión pública basta para imponerse a los innovadores y hacerles guardar silencio y respeto.

Un pasaje de *El Antiguo Régimen y la Revolución* (1856), del mismo Tocqueville, hace una observación similar sobre la religión. En el transcurso de la Revolución francesa,

> quienes mantenían su vieja fe temieron quedar solos en su obediencia y, más espantados por el aislamiento que por la herejía, se unieron a la multitud sin compartir sus creencias. Así, lo que aún era sólo la opinión de una parte de la nación llegó a ser visto como la opinión de todos, y de allí en más pareció irresistible incluso a quienes le habían dado su falsa apariencia.

En estos pasajes, Tocqueville se refiere a las creencias que la gente *profesa* tener (o se abstiene de repudiar), no a las creencias que real y sinceramente sostiene. En este aspecto, su análisis difiere de la conducta en el experimento de la luz en movimiento y en la condición individualmente pública del experimento de concordancia de las líneas.[44] Esta distinción no es, empero, rígida y rápida. Como he sostenido en varios lugares, no siempre está claro qué significa «creer» que algo es como es. Aun en la condición individualmente pública, es posible que la «creencia» de los sujetos que decían que A era la línea concordante haya sido un tanto vaga. Esos participantes quizá no habrían estado dispuestos, por ejemplo, a apostar dinero por esa proposición. Además, la afirmación de una creencia puede, en algunas circunstancias, inducir una tendencia a adherir a ella (Capítulo 7).

La psicología moderna redescubrió la intuición de Tocqueville bajo el encabezado de «ignorancia pluralista». En casos extremos, nadie cree en la verdad de una proposición determinada, pero todos creen que todos los demás creen en ella. En casos más realistas, la mayoría no cree en ella, pero cree que la mayoría sí cree. Ambas situaciones difieren de los casos

patológicos en los cuales todo el mundo profesa públicamente cierta creencia y sabe a la vez que, en realidad, nadie la sostiene en privado. El comunismo desplegó esta *cultura de la hipocresía* en un grado extremo, al menos en su etapa gerontocrática final. La ignorancia pluralista y las culturas de la hipocresía pueden ser alimentadas por el mismo mecanismo, a saber, el temor a la desaprobación o el castigo por enunciar opiniones desviadas. La diferencia es que en la primera, la desaprobación es horizontal: prorrateada por conciudadanos que creen falsamente que deben rehuir a los desviados para no sufrir ellos mismos ese destino. Como señala Tocqueville, quienes no evitan a los desviados pueden ser evitados. En contraste, la cultura de la hipocresía funciona a través de un castigo verticalmente impuesto: quienes no expresen entusiasmo por cumplir el plan u odio hacia el enemigo de clase perderán probablemente su trabajo o algo peor. El castigo vertical tal vez induzca entonces medidas horizontales, si la gente evita o castiga a los desviados para que, a su vez, no la castiguen por desviarse.

La ignorancia pluralista también difiere del mecanismo subyacente al síndrome del espectador pasivo constatado en el asesinato de Kitty Genovese. En (una versión estilizada de) este último caso, cada individuo creía que la pasividad de los otros justificaba la suya. La causa no puede haber sido la presión social o un deseo de ajustarse a las normas grupales, dado que los treinta y ocho espectadores estaban demasiado aislados entre sí para formar una comunidad. Antes bien, la pasividad parecía justificada por una *inferencia*: como al parecer nadie hacía nada, la situación no podía ser muy grave. Esa inferencia se imponía con mucho a los «datos en bruto» (los gritos de la víctima). Dentro de un momento consideraremos con más detenimiento este mecanismo. Aquí, sólo quiero señalar que la situación no implicaba una ignorancia pluralista, dado que no había discrepancias entre lo que cada persona creía en privado y las creencias que atribuía a los demás.

Se ha demostrado que la cultura de la bebida ilustra la ignorancia pluralista. En muchos campus de los Estados Unidos existe una cultura del exceso alcohólico entre los no graduados, sobre todo los varones. La mayoría de los estudiantes se sienten incómodos con los elevados niveles de consumo de alcohol, pero forman parte del proceso porque creen, erróneamente, que casi todos los demás también lo hacen.[45] Su comportamiento en materia alcohólica se ajusta a lo que ellos consideran

equivocadamente como la actitud típica en el campus, y no a sus actitudes privadas. Otro ejemplo puede tomarse de un experimento en el cual se indicó a los estudiantes que leyeran un artículo escrito en un estilo deliberadamente obtuso que lo hacía casi incomprensible, y luego se les preguntó hasta qué punto lo habían entendido y cómo creían que lo habían entendido los otros. En una condición, los estudiantes tenían la opción de acudir al experimentador y pedirle ayuda; en otra, se les dijo en forma expresa que no podían hacerlo. Ni siquiera en la primera condición hubo participantes que fueran a ver al experimentador, porque el procedimiento para recurrir a él implicaba el riesgo de verse en una situación embarazosa. Sin embargo, cada uno de los estudiantes parecía creer que, si bien él no se movilizaba por temor a pasar vergüenza, los demás no lo hacían porque habían entendido el artículo y no necesitaban ayuda. En esa condición, por tanto, los participantes tendían a creer que el resto había entendido el artículo mejor que ellos. La diferencia desapareció en la otra condición. Cabe conjeturar que ese efecto puede deberse a un «síndrome del hermano mayor». Como se ha señalado en el Capítulo 18, todos somos conscientes de nuestras angustias y miedos internos, pero como no tenemos acceso al fuero íntimo de los otros, tendemos a considerarlos más maduros y dueños de sí mismos.

En el estudio de la bebida en los campus, también se verificó que con el transcurso del tiempo, las actitudes privadas, las creencias sobre las actitudes de otros y el comportamiento llegaban a alinearse, lo cual plantea la cuestión de la *estabilidad* de la ignorancia pluralista. En rigor, ésta podría desaparecer de dos maneras: porque las falsas creencias sobre los otros terminan por ser ciertas o porque la gente deja de sostenerlas. Si cada persona adopta la creencia que ha atribuido (falsamente) a los otros, esa atribución llegaría en los hechos a ser cierta. Con toda probabilidad, esto sucedería por obra de la reducción de la disonancia, causada sea por la incomodidad de discrepar con la mayoría o por el disgusto de decir una cosa y creer otra. Al parecer, así sucedió en el tema de la bebida en los campus.

Por otro lado, la situación podría desintegrarse.[46] Supongamos que el 20% de los miembros del grupo muestran en su comportamiento que no tienen la creencia en cuestión, y que el 80% restante la sostiene de la boca para afuera porque requiere más de 20% de no conformistas en el grupo para serlo ellos también. Y supongamos, de manera más específica, que en

un grupo de cien, hay veinte no conformistas, diez que estarían dispuestos a «pronunciarse» si al menos veinticinco ya lo hubieran hecho, quince que lo harían de haber por lo menos treinta y cinco en esa situación, y cincuenta y cinco que se unirían si, como mínimo, cincuenta mostraran su verdadero parecer. Como se ha expuesto, la cultura mayoritaria es estable. Imaginemos, sin embargo, que cinco de los individuos más conformistas se marchan o mueren y son reemplazados por cinco no conformistas. En ese caso, la mayoría se desintegraría. Los veinticinco no conformistas crearían las condiciones para que diez más se unieran a ellos; los treinta y cinco resultantes atraerían a otros quince y se alcanzaría de tal modo el umbral necesario para la incorporación de los cincuenta restantes. En vez de referirse al proceso como *desintegración del conformismo*, también podemos verlo como el *crecimiento del no conformismo como una bola de nieve*. Observaremos una dinámica similar en la acción colectiva (Capítulo 24).

El conformismo puede desintegrarse de muchas otras maneras. El niño pequeño del cuento de Andersen tiene su reflejo en los experimentos de concordancia de las líneas: cuando un *solo* cómplice expuso la opinión veraz de que B era la de longitud más cercana a D, el conformismo casi desapareció. Otro ejemplo: consideremos una creencia difundida tanto en Inglaterra como en Francia con anterioridad a la Reforma, a saber, que el rey podía curar las escrófulas tocando a la persona enferma. La Reforma socavó esa creencia, debido a que los católicos franceses y los anglicanos ingleses se vieron entonces obligados a explicar por qué las pruebas en el otro país eran espurias. Pero el reconocimiento de la posibilidad de un error colectivo de gran escala resultaba peligroso, pues las pruebas presuntamente inválidas utilizadas para respaldar la creencia en el otro país no eran muy diferentes de las que se invocaban en el propio.

Otro mecanismo de desintegración es la publicación de una encuesta de opinión. Con anterioridad al referéndum de 1972 sobre el ingreso de Noruega en el Mercado Común (como se lo denominaba por entonces), el Gobierno, los principales partidos políticos opositores y los grandes diarios estaban masivamente a favor de la entrada. Si bien, según mostró el referéndum, había una mayoría popular contra ella, en el plano individual cada oponente se habría visto inducido a creerse integrante de una pequeña minoría si las encuestas de opinión no hubiesen indicado lo contrario. Sin ellas, el resultado del referéndum habría sido muy probablemente diferente.

Algunos de los que se oponían al ingreso se habrían abstenido de votar, dado que el resultado parecía ineluctable. Por otra parte, el movimiento que se formó para convencer a los indecisos habría seguido siendo pequeño y carente de influencia. En el período transcurrido entre el establecimiento del sufragio universal y el surgimiento de las encuestas de opinión, el campo abierto a la ignorancia pluralista en materia política debe haber sido considerable.

Rumores, temores y esperanzas

Otro cuento de Hans Christian Andersen, «Es la pura verdad», ilustra «la facilidad con que una plumita puede convertirse en cinco gallinas» por medio de sucesivas exageraciones. El estudio de la formación y propagación de los rumores no está, que yo sepa, muy adelantado. Los principales aportes los han hecho los historiadores franceses (y algunos anglo-norteamericanos), que siguen el ejemplo del pionero estudio de Georges Lefebvre sobre el «Gran Pánico» de 1789. Además de ese Gran Pánico, han estudiado los rumores, virtualmente todos falsos, relacionados con:

- El retorno de Napoleón luego de sus derrotas de 1814 y 1815.
- La restauración total del Antiguo Régimen.
- Un reparto socialista de los bienes en vísperas de la Revolución de 1848.
- Una masiva invasión de Alemania en marzo de 1848, encabezada por trabajadores franceses empobrecidos, con saqueos, incendios y asesinatos.
- Una conspiración de médicos para «envenenar al pueblo».
- Una conspiración del clero y la nobleza para «hambrear al pueblo».
- Inminentes reducciones de impuestos.
- Inminentes aumentos de impuestos.
- Francotiradores que disparaban contra los soldados alemanes desde los techos cuando Alemania invadió Bélgica en 1914.
- Decenas de miles de soldados rusos que se unían a las tropas aliadas en agosto de 1914.

Sobre la base de estos estudios es posible, me parece, identificar algunos problemas generales y tal vez extraer ciertas conclusiones. Querría empezar, no obstante, con algunas observaciones empíricas. Primero, la idea de que una pluma se convierte en cinco gallinas no es una exageración del efecto amplificador del rumor. Tras la insurrección de los trabajadores de París en junio de 1848, dos hombres vistos al borde de un camino rural se convirtieron en diez, trescientos y seiscientos en los múltiples relatos

sucesivos, hasta que finalmente pudo escucharse que tres mil «defensores del reparto» (*partageux*) se dedicaban a saquear, quemar y masacrar. Treinta mil soldados fueron enviados a contrarrestar la amenaza. Una investigación reveló que uno de los dos hombres era un loco y el otro era su padre, que lo tenía a su cargo. En esa misma época, un campesino inventó una fantasía para asustar a un niño; poco después, más de mil hombres estaban en armas para derrotar a los inexistentes «bandoleros».

Segundo, a veces es posible identificar con alguna precisión el origen y la velocidad y los mecanismos de propagación de los rumores. El Gran Pánico de 1789 se originó de manera casi simultánea pero independiente (coordinada por la época de la cosecha) en siete lugares diferentes, y luego se difundió (a una velocidad promedio estimada de cuatro kilómetros por hora) por la mayor parte del país. En 1848, el rumor de que trabajadores franceses desempleados iban a invadir Alemania se propagó a una velocidad comparable. En ambos casos, se ha calculado cuánto se reducía la velocidad del rumor al cruzar una montaña o de noche. En muchas circunstancias, las acciones mismas provocadas por él eran, como veremos a la brevedad, su fuente de inspiración. Con frecuencia, el rumor también era propagado por individuos que pretendían, sinceramente o no, estar en posesión de información: funcionarios públicos, vagabundos, buhoneros, pastores, soldados que volvían del frente. Otra fuente de propagación era el tañido de las campanas de las iglesias que podía oírse en las aldeas cercanas. En el período siguiente, los periódicos y las cartas fueron una fuente importante de rumores. Los desmentidos oficiales sobre el meollo de un rumor a menudo servían para alimentarlo.

Lefebvre sintetizó parte de su explicación del Gran Pánico al decir que «el pueblo se temía a sí mismo» (*le peuple se faisait peur à lui-même*). La creencia de que se acercaba un grupo de bandoleros provocaba la movilización de las tropas, y a la distancia se tomaba a otros campesinos por bandidos. Cuando una aldea hacía tañer las campanas de la iglesia, los destacamentos enviados por las aldeas vecinas eran confundidos con enemigos. En 1848, un cañonazo de advertencia lanzado en una aldea francesa se interpretó en las aldeas de las cercanías como el estrépito de la batalla. Cuando los rumores de una inminente invasión de indigentes franceses llegaron a Alemania en marzo de 1848, los trabajadores camineros de la orilla francesa del Rin cruzaron el río a toda prisa para

regresar a sus hogares y sus familias. Otros, al verlos a la distancia, quizás hayan creído que eran los franceses que se aproximaban.

En muchos casos, la sustancia del rumor era la existencia de una conspiración contra el pueblo, orquestada por el gobierno o la elite. Un suceso natural (una mala cosecha, una serie de incendios, un brote de cólera) se atribuía a una agencia deliberada. (Como se ha señalado antes, el pueblo solía considerar que el móvil de las elites era la malevolencia y no el interés.) Con frecuencia, el mismo hábito mental impulsaba a las autoridades a adjudicar una fuente intencional común a lo que eran en sustancia rumores surgidos en forma independiente. Del hecho de que rumores similares aparecieran de manera simultánea en diferentes partes del país, las autoridades inferían correctamente que debían tener una causa común. En vez de identificar esa causa como una situación objetiva compartida, una mala cosecha, por ejemplo, deducían incorrectamente que había un agente intencional en acción. Los rumores sobre conspiraciones eran inseparables de la creencia de que las conspiraciones creaban rumores.

En un comentario de la tendencia al crecimiento de los rumores, Montaigne planteó lo que tal vez sea el primer análisis de los micromecanismos de su transmisión:

> Hay más distancia entre nada y la cosa más pequeña del mundo, que entre ésta y la más grande. Y resulta que los primeros que se abrevan con este comienzo de rareza, al ir sembrando su historia, van sintiendo por las objeciones que se les hacen, dónde reside la dificultad de persuadir, y van calafateando ese lugar con alguna pieza falsa. […] El error individual provoca primero el error público, y después, a su vez, el error público provoca el error particular. Así va toda esta obra, construyéndose y formándose de mano en mano; de manera que el testigo más alejado está más enterado que el más próximo, y el que se ha informado el último, más convencido que el primero. Es una evolución natural. Pues todo el que cree algo considera que es un deber de caridad el convencer de ello a otro; y para hacerlo, no teme añadir todo cuanto crea necesario para su historia, para vencer la resistencia.

La descripción es un tanto benévola, pues Montaigne no atribuye al propagador de rumores otra cosa que el deseo de persuadir a otros de lo que él cree cierto. En su análisis del Gran Pánico, Lefebvre menciona otros motivos, más oscuros. Algunos individuos podían tener motivos para exagerar el peligro a fin de que no se los acusara de cobardía. La incredulidad manifiesta podía suscitar a la vez acusaciones de actuar al servicio de la contrarrevolución, pues de ese modo se adormecía al pueblo en medio del peligro, y el riesgo de ofender el amor propio de quienes activaban la alarma. En algunos análisis de los rumores en el siglo XIX se ha

argumentado que los vagabundos tendían naturalmente a propagar aquellos que sus oyentes querían escuchar (una forma de autoengaño inducido): el retorno de Napoleón I para complacer a sus seguidores o la enfermedad de Napoleón III para complacer a sus adversarios. De manera similar, los buhoneros contaban los rumores más sensacionalistas que pudieran atraer una vasta audiencia. En lo concerniente a los rumores sobre los francotiradores belgas, su exactitud parecía incontrovertible una vez que se los utilizaba como premisa de represalias sangrientas. ¿De qué otra manera, habrán pensado quizá los alemanes, podían justificarse sus atrocidades? Y cuando los rumores eran transmitidos por soldados heridos enviados a sus casas desde el frente, ¿quién se habría atrevido a contradecirlos?

Los autores que estudian los rumores hacen hincapié en que éstos surgen de la «conciencia colectiva», bajo la forma de esquemas preexistentes que aun sucesos insignificantes pueden activar. Las creencias populares acerca de la malevolencia de las elites y su tendencia a montar conspiraciones contra el pueblo se cuentan entre las condiciones generales que hacen creíbles los rumores, además de los factores específicos ya enumerados. Las creencias sobre la insaciable necesidad gubernamental de mayores impuestos y más soldados también facilitaban que acciones aparentemente neutras como las investigaciones con fines estadísticos generaran rumores acerca de la inminencia de aumentos impositivos y conscripciones masivas. En 1914, las creencias alemanas concernientes a la naturaleza de la resistencia belga se formaron en virtud de la activación de esquemas creados durante la guerra franco prusiana de 1870-1871 y referidos a los francotiradores. Por lo común, esos esquemas tenían cierto fundamento en la realidad, aun cuando las creencias específicas que inspiraban carecieran de toda base.

He citado varias veces el proverbio «creemos con facilidad lo que esperamos y lo que tememos». Los rumores que he examinado ilustran ambas posibilidades. El retorno de Napoleón, una baja de los impuestos en oportunidad de un cambio de régimen y la adhesión de los soldados rusos a las fuerzas aliadas dan testimonio del poder de la esperanza. (En el caso de Napoleón, sin embargo, también había quienes temían ese retorno.) Pese a ello, un historiador de los rumores en la Francia decimonónica afirma que, «en general, los rumores son sistemáticamente mucho más pesimistas que eufóricos»: el miedo prevalece sobre la esperanza. Para verificar esta afirmación, podríamos tratar de enumerar episodios de pensamiento

desiderativo y contradesiderativo, para ver si, en efecto, el segundo predomina sobre el primero. Probablemente ésta sería, sin embargo, una empresa desesperada, tanto por la imposibilidad de establecer una muestra representativa como por la dificultad de saber cuánta importancia asignar a cada episodio.

Un método más prometedor consiste en basarse en la distinción entre creencias y cuasicreencias que propuse en la introducción a la segunda parte; la principal diferencia es que sólo las primeras se utilizan como premisas de la acción. Si consideramos episodios de formación de rumores, al parecer se constata casi sin excepción que *sólo los basados en el miedo llevan a la gente a modificar su comportamiento*. En los episodios de formación de rumores en la Francia de los siglos XVIII y XIX, los campesinos cosechaban el grano antes de que estuviera maduro para impedir que los «bandoleros» lo destruyeran, vendían su cosecha sin guardarse siquiera semillas porque temían que todo fuera confiscado, escondían sus objetos de valor cuando se rumoreaba que tendrían que pagar impuestos por ellos, se casaban para evitar la conscripción y acaparaban sal cuando circulaba el rumor de que la afectaría un nuevo gravamen. Los rumores sobre la aproximación de bandoleros tuvieron una influencia decisiva en el rumbo de la Revolución francesa. Entre otras cosas, llevaban a los campesinos a atacar los castillos de sus señores, acciones que a su turno provocaron los decretos del 4 de agosto de 1789, de abolición del feudalismo. De manera análoga, los rumores sobre el ataque inminente de otro grupo suscitan con frecuencia disturbios étnicos.[47] El pensamiento contradesirativo generado por el miedo parece tener gran capacidad de influencia sobre el comportamiento. En contraste, la adhesión al pensamiento desiderativo generado por la esperanza parece deberse más a su valor de consuelo. En apariencia, esta asimetría es muy sólida. Si bien no prueba que los rumores basados en el miedo sean más numerosos, sí muestra que tienen mayor importancia para la explicación del comportamiento.

La principal excepción a la asimetría con que me he topado se produce en los mercados financieros, donde los rumores alimentados por la esperanza se basan a menudo en burbujas especulativas. La «exuberancia irracional» de la década de 1990 se utilizó sin duda como base para la acción, una acción que a la larga fue contraproducente. Al mismo tiempo, los mercados financieros también exhiben la influencia del pesimismo irracional. Sin embargo, poco parece conocerse sobre la mecánica y la

dinámica de la formación *interactiva* de creencias en esos mercados. Hay al parecer una compleja interacción entre quienes actúan sobre la base de rumores, quienes actúan fundados en los cambios de precios causados por los rumores y quienes actúan en función de rumores cuando éstos cuentan con el respaldo de los cambios de precios. Aunque algunos agentes actúen en función de creencias que, dadas las pruebas existentes, son racionales, dichas pruebas son en parte el resultado de acciones irracionales llevadas a cabo sobre la base de rumores infundados.

La asimetría parece limitarse a la formación interactiva de creencias. En el nivel del individuo, el pensamiento desiderativo es, a no dudar, tan capaz de generar comportamientos como el pensamiento contradesiderativo. Por otra parte, la observación casual sugiere que en el plano individual el último mecanismo es menos corriente que su opuesto. La naturaleza interactiva de los rumores parece generar, de una manera que no comprendemos del todo, patrones que se apartan de los que observamos cuando los individuos forman sus creencias sin contacto con los demás.

Cascadas informacionales

Los rumores también pueden surgir a raíz de una formación de creencias totalmente racional, por medio de un mecanismo conocido como «cascadas informacionales». Supongamos que cada individuo de un grupo tiene acceso a cierta información privada sobre algún tema. Todos forman sus creencias de manera secuencial, cada uno apoyado en su información privada *y* en las creencias expresadas por sus predecesores (si los hay) en la secuencia. Cada aldeano, por ejemplo, podría tener algunas pruebas privadas acerca de la presencia de bandoleros en la vecindad y utilizarlas, junto con lo que ha oído de otros, para formarse la opinión que luego transmite. La votación por lista puede tener la misma dinámica, si el asunto en cuestión sólo gira en torno de creencias sobre problemas fácticos y no de preferencias. Cada integrante de una asamblea se apoyará no sólo en su propia información sino también en lo que revela el voto de quienes lo han precedido en la lista. Un tercer ejemplo podría ser el del árbitro de una revista que se entera de que el artículo que debe juzgar ha sido rechazado por el árbitro de otra publicación (aunque no sabe por qué).

En estas situaciones, la gente usa las conclusiones del proceso de formación de creencias de otros como insumos indirectos para su propia

formación, sin conocer los insumos directos (la información privada) que esos otros utilizaron para llegar a sus conclusiones. Puede suceder entonces que individuos racionales terminen por abrigar creencias falsas, aunque hubieran llegado a la conclusión acertada de haber tenido cada uno de ellos acceso a los «datos en bruto» de sus predecesores, y no sólo a sus conclusiones. En el ejemplo de los árbitros, el segundo lector, si hubiera leído el primer informe, podría haber detectado una parcialidad o un razonamiento defectuoso. Sin embargo, si sólo conoce la conclusión de ese primer informe y la excelente reputación de que disfruta la revista correspondiente, es racional que tome en cuenta la opinión negativa del otro árbitro, junto con su propia evaluación. Si ésta es favorable, pero sólo lo es hasta cierto punto, quizá termine por recomendar el rechazo. Un tercer revisor con una opinión personal muy favorable también podría (racionalmente) aconsejar el rechazo si se entera de que los dos primeros árbitros han hecho otro tanto. No obstante, es posible que el resultado sea menos que óptimo (en relación con los objetivos de la comunidad académica), dado que el segundo y el tercer árbitros han favorecido la publicación y el primero quizás haya expresado sólo una tibia oposición a ella.[48] Si los revisores hubiesen leído el artículo en el orden inverso, la conclusión habría sido diferente («dependencia del camino elegido»).

* * *

NOTA BIBLIOGRÁFICA

Los experimentos de concordancia de las líneas, llevados a cabo por primera vez por Solomon Asch, se describen en cualquier manual de psicología social, por ejemplo Elliot Aronson, *The Social Animal*, 9ª edición, Nueva York, Freeman, 2003 [*El animal social*, Madrid, Alianza, 2000]. El experimento de la luz en movimiento se describe en Robert C. Jacobs y Donald T. Campbell, «The perpetuation of an arbitrary tradition through several generations of laboratory microculture», *Journal of Abnormal and Social Psychology*, 62(3), 1961, págs. 649-658. Sobre el consumo de alcohol en los campus, véase Deborah A. Prentice y Dale T. Miller, «Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm», *Journal of Personality and Social Psychology*, 64(2), 1993, págs. 243-256. Con respecto a los estudiantes expuestos a la lectura del artículo incomprensible, véase Dale T.

Miller y Cathy McFarland, «Pluralistic ignorance: when similarity is interpreted as dissimilarity», *Journal of Personality and Social Psychology*, 53(2), 1987, págs. 298-305. El escenario de la desintegración se apoya en Timur Kuran, *Private Truths, Public Lies: The Social Consequences of Preference Falsification*, Cambridge (Massachusetts), Harvard University Press, 1995. La observación acerca del efecto de la Reforma sobre la creencia en el poder de curar del rey se debe a Marc Bloch, *Les Rois thaumaturges: étude sur le caractère surnaturel attribué à la puissance royale particulièrement en France et en Angleterre*, París, Armand Colin, 1961 [*Los reyes taumaturgos*, México, Fondo de Cultura Económica, 1988]. Los estudios de la formación de rumores en los que me he basado son Georges Lefebvre, *La Grande peur de 1789*, París, Armand Colin, 1988 [*El gran pánico de 1789: la Revolución francesa y los campesinos*, Barcelona, Paidós, 1986]; François Ploux, *De bouche à oreille: naissance et propagation des rumeurs dans la France du XIX^e siècle*, París, Aubier, 2003; Ralph C. Cenevali, «The "false French alarm": revolutionary panic en Beden, 1848», *Central European History*, 18(2), junio de 1985, págs. 119-142; Marc Bloch, «Réflexions d'un historien sur les fausses nouvelles de guerre», *Revue de synthèse historique*, 33, 1921, págs. 13-35, y Christophe Prochasson y Anne Rasmussen (comps.), *Vrai et faux dans la Grande guerre*, París, La Découverte, 2004. Se encontrarán un catálogo y un análisis detallados de los rumores en los disturbios étnicos en Donald Horowitz, *The Deadly Ethnic Riot*, Berkeley, University of California Press, 2001. Sobre los rumores en el mercado bursátil, véase Arnold M. Rose, «Rumor in the stock market», *Public Opinion Quarterly*, 15, 1951, págs. 61-86. Sushil Bikchandani, David Hirshleifer e Ivo Welch, «Learning from the behavior of others: conformity, fads, and informational cascades», *Journal of Economic Perspectives*, 12(3), verano de 1998, págs. 151-170, es una introducción al mecanismo de las cascadas informacionales.

Notas al pie

[42] En el Capítulo 5 examiné el hecho de que la observación de otros puede suscitar un comportamiento similar al suyo por obra de la norma cuasimoral de la equidad, mientras que el hecho de ser observado *por* otros puede provocar un comportamiento similar en virtud del miedo a la desaprobación. En la formación de creencias, la observación por parte de otros también puede generar conformidad por obra del miedo a la desaprobación, mientras que la conformidad producida cuando se observa a otros puede ser el resultado del aprendizaje o de la reducción de la disonancia.

[43] En el ejemplo numérico utilizado para ilustrar el aprendizaje bayesiano en el Capítulo 11, cada nuevo elemento de prueba confirmatorio provoca un aumento más pequeño que el anterior en la probabilidad. Este «valor marginal decreciente de la nueva información» es un fenómeno bastante general.

[44] En realidad, el experimento de la luz en movimiento era doblemente público y dejaba campo a la insinceridad. Sin embargo, es de presumir que la naturaleza ambigua de la tarea facilitaba la adopción sincera, o casi sincera, de la creencia exagerada. Supongo que la razón del procedimiento era que el estudio de generaciones sucesivas habría sido difícil de hacer en una condición individualmente pública.

[45] Acaso sea cierto al mismo tiempo, sin embargo, que la mayoría de los estudiantes no beben y que la mayoría de los amigos de la mayoría de los estudiantes beben, si quienes beben tienen más amigos que quienes no beben.

[46] La desaparición de las falsas creencias en el experimento de la luz en movimiento también es una forma de desintegración, debido al hecho de que en cada generación los sujetos utilizan sus propios «datos en bruto» para ajustar la distancia estimada un tanto hacia abajo, en comparación con lo que escuchan de los otros.

[47] El papel de la emoción en los disturbios es confirmado por su tendencia documentada a ocurrir más frecuentemente cuando hace calor.

[48] Varias prácticas que reducen la conformidad pueden entenderse en esta perspectiva. Cuando (como solía suceder en Noruega) el calificador interno de los exámenes universitarios envía la respuesta del estudiante a un calificador externo, no le comunica su propia nota. Por la misma razón, cuando uno busca una segunda opinión médica, no debe decir al segundo médico qué ha dicho el primero.

Capítulo 24

Acción colectiva

La Bandera Viviente

En *Lake Wobegon Days*, Garrison Keillor describe el Día de la Bandera en su pueblo mítico. Herman, el organizador del desfile, solía comprar una cantidad de gorras azules, rojas y blancas y las repartía entre los residentes del pueblo, para que éstos pudieran marchar por las calles como una bandera viviente, mientras él, desde el techo del Edificio Central, les tomaba una fotografía. Apenas terminada la guerra, la gente aceptaba con entusiasmo la tarea, pero más adelante comenzó a pensarlo dos veces:

> Una de las causas de enfado era el hecho de que ninguno de ellos lograra ver la bandera que formaban; en el diario, la foto se publicaba en blanco y negro. Sólo Herman y el señor Hanson llegaban a ver la bandera real, y algunos niños demasiado pequeños como para ser necesarios allá abajo. La gente quería tener la posibilidad de subir al techo y presenciar el espectáculo por sí misma.
>
> «¿Cómo vais a subir si debéis estar aquí abajo?», decía Herman. «Si subís a mirar, no habrá nada para ver. ¿No basta con saber que estáis haciendo vuestra parte?»
>
> El Día de la Bandera de 1949, justo en el momento en que Herman decía: «¡Eso es! ¡Ahora mantenedla así!», uno de los rojos se escabulló; subió a la carrera cuatro tramos de escalera hasta el techo, se asomó y echó una larga mirada. A pesar del hueco que había dejado, la vista era magnífica. Allá abajo, la Bandera Viviente se extendía a lo largo de tres calles. ¡Una bandera perfecta! ¡Esos rojos, tan brillantes! No podía sacarle los ojos de encima. «¡Baje de aquí! ¡Tenemos que tomar una foto!», le ladró Herman. «¿Cómo está?», le preguntaron a gritos desde abajo. «¡Es increíble! No puedo describirla», contestó el rojo.
>
> De modo que todos quisieron mirar. «¡No!», dijo Herman, pero se votó y el resultado fue unánime. Uno por uno, los miembros de la Bandera Viviente subieron al techo y la admiraron. *Era* maravillosa. Los ojos se llenaban de lágrimas, y la imagen hacía reflexionar sobre este gran país y el lugar que Lake Wobegon ocupaba en él. Uno quería quedarse allí arriba toda la tarde y

no hacer otra cosa que embeberse en ella. Así, transcurrida la primera hora y cuando apenas cuarenta de los quinientos habían estado en el techo, los otros comenzaron a impacientarse cada vez más. «¡Apresuraos! ¡Dejad de perder el tiempo! ¡*Ya* la habéis visto! ¡Bajad y dejad que otros tengan la oportunidad!» Herman decidió enviar grupos de cuatro personas y luego de diez, pero al cabo de dos horas la Bandera Viviente se había convertido en la Bandera Sentada y luego empezó a deshilacharse, ya que las personas que habían subido sólo pensaban en volver a casa a cenar, lo cual enfurecía a quienes no habían estado arriba.

«¡Diez minutos más!», exclamó Herman, pero los diez minutos se hicieron veinte y treinta, la gente comenzó a irse furtivamente y la bandera que se ofreció a la vista del último espectador era una bandera desgarrada por fuego de cañones.

En 1959, Los Hijos de Knute tomaron a su cargo el Día de la Bandera. Herman les entregó las bolsas de gorras. Desde entonces, los Knutes han conseguido armar varias buenas banderas, aunque la mayor parte de los años la concurrencia ha sido escasa. Hacen falta al menos cuatrocientas personas para hacer una buena. Algunos años, los Knutes prohibían subir a mirar desde el techo, otros organizaban una lotería. Una vez experimentaron con un gran espejo sostenido por dos hombres sobre el borde del techo, pero cuando la gente se retrepaba y alzaba la vista, la bandera, por supuesto, desaparecía.[49]

Los habitantes de Lake Wobegon enfrentan un problema de acción colectiva, si bien bastante desusado. Cada uno de ellos está tentado de subir al techo o, en el experimento posterior, de mirarse en el espejo. Pero si todos lo hacen, la bandera se *desintegra* o desaparece y no hay nada que mirar. Para resolver el problema, recurren a técnicas clásicas de coordinación: imponer la prohibición de «mirar» para impedir que haya aprovechados, subir por turno, hacer un sorteo. La primera no es muy satisfactoria, dado que no tiene mucho sentido formar una bandera que nadie puede ver.[50] Las otras garantizan que algunas personas puedan mirar, pero no tantas como para que ya no quede nada para ver. Exigen, no obstante, una coordinación centralizada, respaldada por organizaciones y sanciones. El desafío es entender de qué manera la gente resuelve a veces sus problemas de acción colectiva mediante una acción descentralizada.

Desintegración y bola de nieve

Si la gente quedara librada a sus propios recursos, la bandera se *desintegraría* literalmente. En un sentido metafórico, la desintegración también se ha mostrado en experimentos con bienes públicos. Los sujetos reciben determinada suma de dinero y se les dice que si aportan una parte a un pozo común, éste se multiplicará, y el monto incrementado se distribuirá luego en partes iguales entre todos los participantes, hayan o no hayan hecho un aporte. (Este es un DP de muchas personas, jugado una serie de

veces por los mismos sujetos.) Si todos los sujetos fueran racionales y egoístas, nadie aportaría nada. De hecho, aun en las interacciones anónimas los sujetos comienzan por aportar, en promedio, entre el 40% y el 60% de su dotación. En rondas posteriores, los aportes caen ininterrumpidamente hasta estabilizarse en torno del 10%. Una posible explicación es que la mayoría de la gente está dispuesta a contribuir si y sólo si los demás hacen cierta contribución. El nivel de ésta puede variar según los individuos, de modo que algunos exigen un elevado aporte promedio para decidirse a hacer su propia donación, mientras que otros aportarán siempre que algunos otros contribuyan, e incluso si nadie lo hace. La desintegración puede suceder entonces como lo muestra la Figura 24.1.

En el comienzo, la gente cree que los otros aportarán, en promedio, OD. Algunos individuos son completamente egoístas y no aportan nada, pero otros, movidos por las normas de equidad, deciden contribuir. La suma total de sus contribuciones asciende a OH. Este monto más pequeño (OH = OC) se utiliza luego como base para el monto esperado en el siguiente juego. Algunos de los que estaban dispuestos a dar cuando los aportes esperados fueran OD, ahora abandonan, de modo que las contribuciones resultantes caen a OG. En rondas subsiguientes, más y más aportantes abandonan cuando el nivel de aportes cae por debajo del nivel en que están decididos a contribuir, hasta que el proceso se estabiliza en OA = OE. En este punto, las contribuciones reales y esperadas coinciden.

También puede observarse el fenómeno opuesto, la «bola de nieve», en lugar de la «desintegración». Un ejemplo fue la acumulación de multitudes a lo largo de domingos sucesivos en Leipzig antes de la demolición del Muro de Berlín en 1989. También en las revoluciones y rebeliones, un pequeño número de participantes iniciales puede crecer hasta transformarse en un movimiento que derroca el Gobierno. Sin embargo, antes de procurar entender fenómenos de esas características, es preciso exponer con más detenimiento la estructura de los problemas de la acción colectiva. Por ahora, podemos abstraernos del aspecto secuencial de los ejemplos recién dados y definir un problema de acción colectiva simplemente en términos de la relación entre elecciones individuales y resultados colectivos.[51]

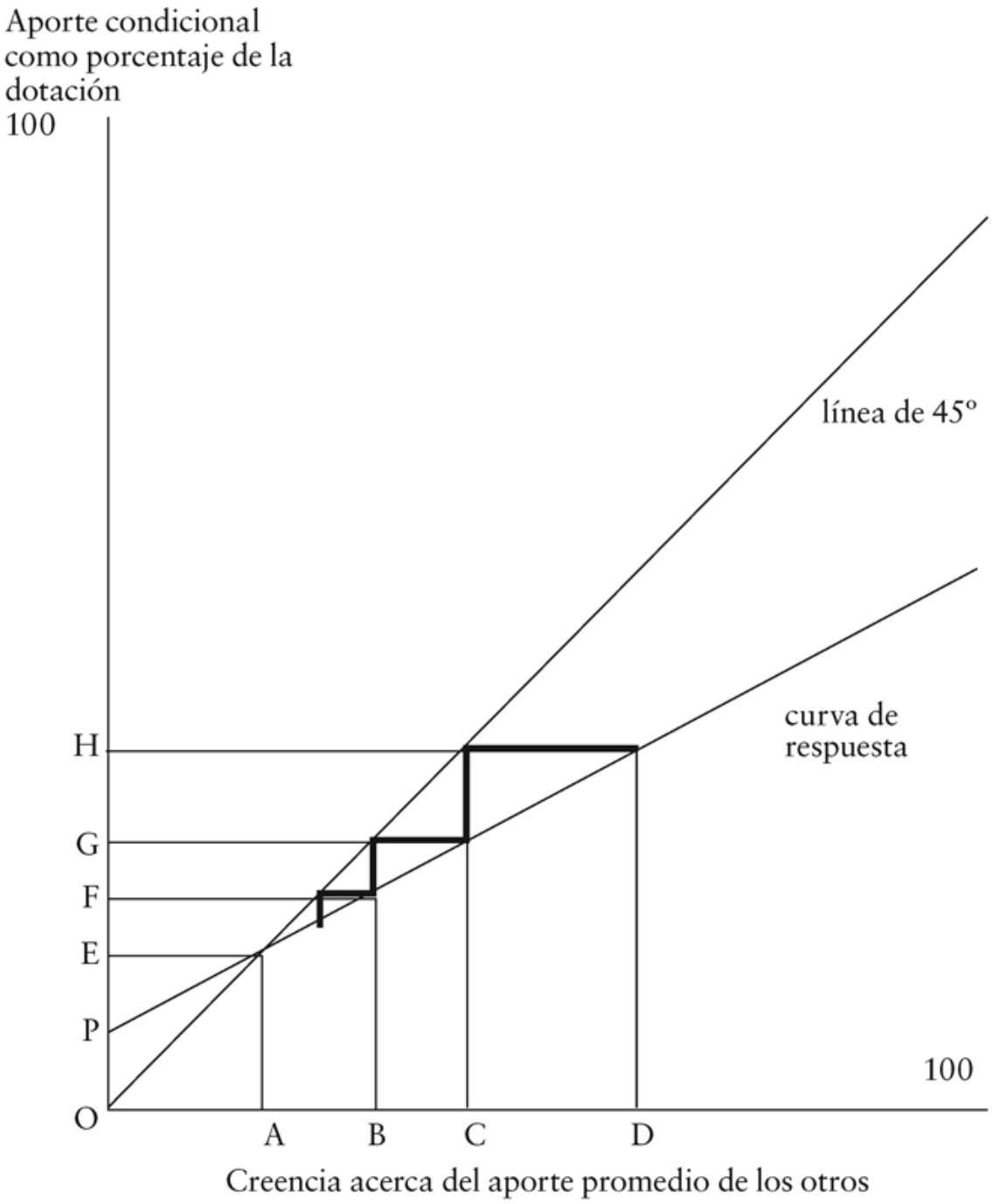


FIGURA 24.1

El problema está estrechamente relacionado con las cuestiones que he examinado en capítulos anteriores. Los problemas de acción colectiva surgen en presencia de externalidades negativas o positivas que generan un

DP de muchas personas o una situación conexa. Para *definir* un problema de acción colectiva, me limitaré a un subconjunto de las motivaciones que pueden animar a los agentes en tales situaciones. Específicamente, se supone que éstos sólo están motivados por los costes en que incurren al participar y por los beneficios personales que extraen del resultado de la acción colectiva. Un trabajador en huelga, por ejemplo, debe tomar en cuenta únicamente el riesgo de perder el trabajo o el salario mientras dure la medida de fuerza, y la perspectiva de un salario más alto (para él) si ésta triunfa. En contraste, tal vez tengamos que apelar a un conjunto más amplio de motivaciones para *explicar* la elección de la estrategia cooperativa en estas situaciones.

La provisión de una externalidad positiva tiene costes o riesgos directos para el agente. En el experimento de los bienes públicos, el coste consiste en ceder parte de la propia dotación. En la marcha celebrada en Leipzig, era el de ser molido a palos por la policía. Al mismo tiempo, el agente proporciona un pequeño beneficio material a las personas con las cuales interactúa. En el experimento, cada uno de los demás sujetos recibía una fracción de su aporte. También el sujeto obtenía un beneficio material (se le entregaba la misma fracción), pero obviamente menor que su coste. En Leipzig, cada manifestante representaba un punto de mira adicional para la atención de la fuerza policial (que se suponía constante) y, de ese modo, hacía marginalmente menos probable que cualquier manifestante en particular recibiera una paliza. Así, si nos limitamos a los costes y beneficios directos de la cooperación, la defección individual supera la cooperación individual. Al mismo tiempo, en ambos casos la cooperación universal supera la defección universal. Si todos aportan su dotación entera al pozo común, todos reciben como reembolso el mismo múltiplo. Si todos los habitantes de Leipzig salen a las calles, el riesgo de que cualquiera de ellos sea molido a palos se acerca a cero, mientras que las posibilidades de que el régimen odiado caiga se aproximan a uno.

La tecnología de la acción colectiva

Podemos mostrar esta situación mediante un diagrama. En un grupo de $n + 1$ individuos, la Figura 24.2 indica cuánto varía la retribución dada a un individuo determinado como función de su comportamiento y el de los otros. El comportamiento de estos últimos se indica a lo largo del eje

horizontal, que mide la cantidad de cooperadores (entre los otros). Si el individuo también es un cooperador, su utilidad, medida a lo largo del eje vertical, se indica en la línea R AB de los diagramas. Si es no cooperador, su utilidad se mide a lo largo de la línea L OC. Las líneas L y R intersecan los ejes verticales en el orden que define el DP corriente (de dos personas): el resultado preferido es la no cooperación unilateral (ser un aprovechado); el siguiente es la cooperación universal; el tercero, la no cooperación universal, y el peor, la cooperación unilateral (ser explotado). Como en el caso de las dos personas, la no cooperación es una estrategia dominante, puesto que la línea L está en todos los puntos por encima de la línea R. Sin embargo, en contraste con el caso de las dos personas, podemos definir un número M de cooperadores que son capaces de alcanzar una mejor situación si cooperan, incluso en presencia de aprovechados a quienes ponen aun en mejor situación. La línea OB muestra el beneficio *promedio* para todo el mundo, cooperadores y no cooperadores, como función del número de cooperadores. Como el número de agentes es constante, OB también reflejará el beneficio *total* producido por la cooperación.

La situación de la Figura 24.2 refleja un caso especial. Supone que el coste de la cooperación, medido por la distancia entre las curvas L y R, es constante. En otros casos, ese coste aumenta a medida que más personas cooperan. Cuando la gente participa en las campañas de llamadas en defensa de la radio pública, las líneas se congestionan y las comunicaciones se demoran. Puede suceder entonces que los últimos en sumarse[52] reduzcan en realidad el beneficio promedio, porque, para ellos, el coste de participar supera la suma de los beneficios que generan para todo el resto (y para sí mismos). El coste también puede ser inicialmente alto y luego decrecer. A medida que más personas se sumaban a las marchas en Leipzig, las fuerzas del Gobierno debían adoptar un orden cada vez más disperso. También en ese caso, y por una razón similar, los primeros cooperadores tal vez hayan empeorado realmente la situación (en el sentido de reducir el beneficio promedio).

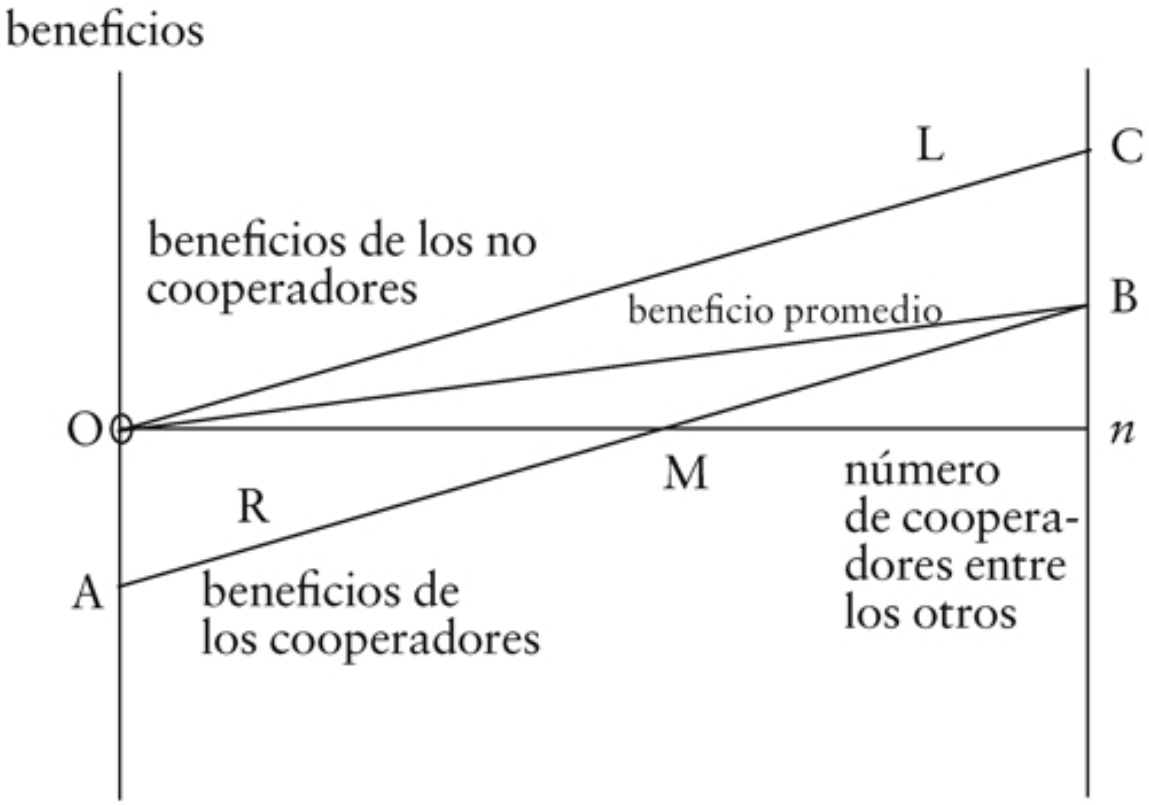


FIGURA 24.2. Tomada con modificaciones de Thomas C. Schelling, *Micromotives and Macrobehavior*, Nueva York, Norton, 1978 [*Micromotivos y macroconducta*, México, Fondo de Cultura Económica, 1989].

La Figura 24.2 supone asimismo que los beneficios de la cooperación, dados por la línea L, son una función lineal del número de cooperadores. Cada nuevo cooperador agrega un aporte de la misma magnitud al bienestar de todos. Los beneficios marginales crecientes pueden ilustrarse con la limpieza de basura de una playa: la última botella que se saca contribuye más que la penúltima a marcar una diferencia estética. Los beneficios marginales decrecientes también son frecuentes. Un ejemplo simple es el de la llamada a la alcaldía para reclamar debido a la presencia de un bache en una zona urbana de clase media: la primera persona que se toma el tiempo de llamar podría hacer que la probabilidad de reparación del bache fuera de 0,4; la segunda la elevaría a 0,7; la tercera, a 0,8; la cuarta, a 0,85; la quinta, a 0,88, y así sucesivamente. A veces, tanto los primeros como los últimos aportantes agregan poco, mientras que los ubicados en el medio son más

eficaces. Algunos revolucionarios o huelguistas no hacen mucha diferencia, y cuando casi todo el mundo se ha sumado importa poco si los escasos no comprometidos también lo hacen. En los movimientos sociales, este patrón es probablemente típico.

Los beneficios marginales de la cooperación tal vez sean incluso negativos en cierto espectro de cooperadores. El desarme unilateral puede hacer que todas las naciones estén peor si genera un vacío de poder que es preciso llenar, y desencadena así una guerra general. Los actos aislados de rebelión pueden dar a las autoridades un pretexto para tomar medidas enérgicas no sólo contra los rebeldes reales, sino también contra los rebeldes potenciales. A la inversa, quizás haya demasiados cooperadores. Supongamos que en tiempos de guerra todo el mundo insiste en alistarse en el ejército, de manera tal que las industrias vitales para el esfuerzo bélico se quedan sin personal suficiente y la guerra se pierde. Si todos insisten en ayudar con la cena en la excursión, muchas manos en un plato harán mucho garabato.

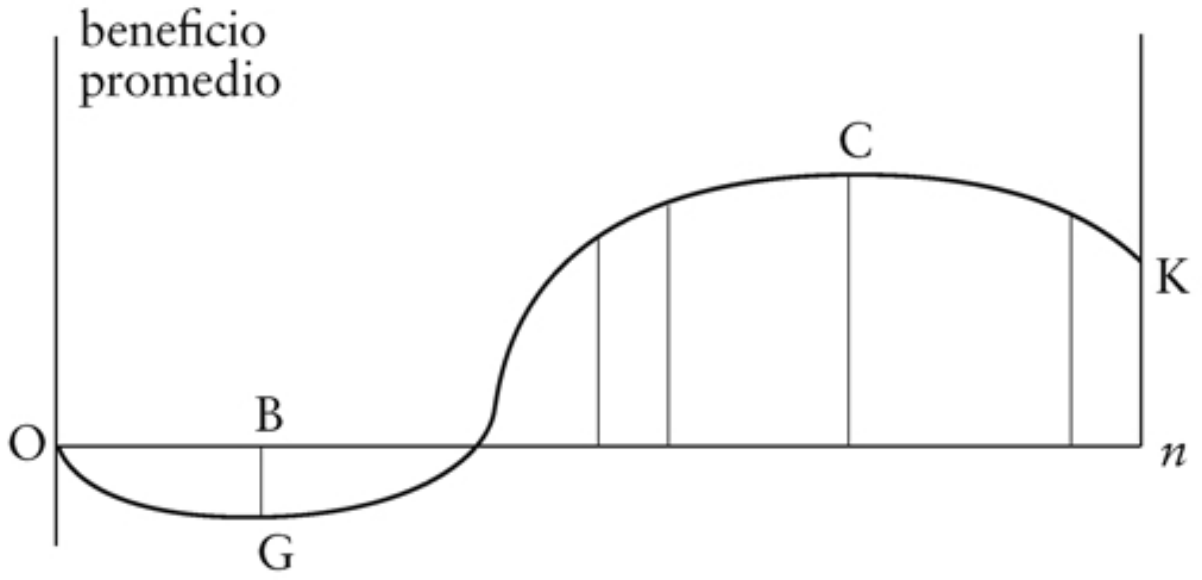


FIGURA 24.3

Como muestran estas observaciones, la tecnología de la acción colectiva difiere de caso en caso. En lo que sigue, me concentro en el caso mostrado en la Figura 24.3, que me parece bastante característico de los movimientos

sociales que tratan de provocar un cambio de políticas. Los primeros colaboradores incurren en altos costes o riesgos y producen pocos beneficios para los otros. De hecho, tal vez perjudiquen a éstos en lugar de beneficiarlos. Su contribución neta es negativa. Los últimos colaboradores también producen escasos beneficios. En algunas circunstancias, su coste puede ser decreciente, como he sugerido. En otras, todos los que luchan por una causa pueden llegar a incurrir en costes o riesgos considerables hasta que el adversario capitula. Quienes se sumaron a la Resistencia francesa en 1944 a menudo infligieron poco daño a los alemanes, pero en todos los casos su vida corrió un peligro sustancial.

La superación del problema de los aprovechados

He definido el problema de la acción colectiva en términos de *motivaciones egoístas racionales orientadas a los resultados*. Por lo común, éstas son insuficientes para suscitar la cooperación, dado que, hagan los otros lo que hicieren, el individuo estará mejor si no coopera. Querría mencionar dos excepciones a este enunciado.

La primera se plantea si los *mismos* agentes se enfrentan *repetidas veces* al mismo problema de acción colectiva. En una interacción continua de este tipo, una *estrategia de gatillo* puede sostener la cooperación: cooperar mientras todos los demás lo hagan, y desertar para siempre si un solo agente deserta una vez.[53] Hay ciertas pruebas de que los cárteles, por ejemplo, se mantienen de ese modo. Las decisiones de no recurrir al uso de gases en la Segunda Guerra Mundial también pueden haberse debido al temor a las represalias, pero más probablemente a deficiencias operativas del método. Sin embargo, estos casos no son típicos. Los trabajadores, por ejemplo, no deciden todos los años si van a renovar o no su afiliación sindical. Las rebeliones y revoluciones son movimientos únicos a los que el argumento tomado de la interacción repetida no se aplica. Otras campañas, cruzadas e impulsos suelen tener demasiado poca estructura para que los cooperadores identifiquen a los no cooperadores. Para explicar en tales casos cómo surge la cooperación, debemos buscar en otra parte.

La segunda excepción se plantea si hay un agente exterior, por lo común una organización, que puede imponer un castigo selectivo a los no cooperadores u otorgar recompensas selectivas a los cooperadores. El Estado quizá castigue a los individuos que imponen externalidades

negativas a otros, por ejemplo cuando fuman en público o escupen en la calle. Los sindicatos tal vez ofrezcan vacaciones baratas a sus afiliados. Los movimientos revolucionarios pueden prometer a los activistas cargos de importancia en el Gobierno posrevolucionario. Estas «soluciones» al problema de la acción colectiva se apoyan en la imposición centralizada, no en la acción descentralizada de los individuos directamente involucrados. Nótese, no obstante, que la existencia misma de una organización capaz de modificar los incentivos de los agentes sociales puede ser el resultado de una acción colectiva exitosa. Además, las promesas de recompensar a los activistas tras la victoria quizá no sean creíbles. En muchos casos, desde Enrique IV en 1594 hasta Charles de Gaulle en 1945, líderes carismáticos han decepcionado a sus seguidores al designar o mantener a funcionarios del régimen derrotado en cargos de primera línea. Los militantes no siempre son buenos administradores.

En las situaciones únicas y descentralizadas de acción colectiva, no habrá cooperación si todos los agentes son plenamente racionales, sólo tienen motivaciones egoístas orientadas a los resultados y estos dos hechos son de conocimiento general. (Esta proposición es una verdad lógica, no una afirmación empírica.) Con referencia a la Figura 24.3, podemos distinguir una serie de motivaciones capaces de suscitar y sostener un comportamiento cooperativo. En el escenario estilizado que bosquejaré, esas motivaciones hacen su aporte sucesivamente, apoyada cada una de ellas (excepto la primera) en la presencia de agentes motivados por los agentes precedentes. Aunque este escenario específico no se aplica a todos los casos, creo que una de sus características tiene validez universal. La acción colectiva exitosa, cuando la observamos, no es obra de individuos con motivaciones idénticas. Antes bien, es el resultado de *motivaciones mixtas*, tanto en el plano individual como en el conjunto de los individuos.[54] A continuación identificaré media docena de dichas motivaciones, que, a mi entender, están presentes en diversas proporciones en muchas situaciones de acción colectiva.

La cooperación requiere la actuación de algunos individuos que no estén meramente motivados por los costes y beneficios para sí mismos. Dos categorías de esos individuos son los que llamaré *utilitaristas plenos* y *utilitaristas desinteresados*. Los primeros cooperarán si y sólo si su contribución incrementa el beneficio promedio. En la Figura 24.3, requieren que el número de cooperadores ya haya llegado a un punto OB donde su

acción tiene consecuencias útiles. La segunda categoría cooperará si y sólo si su contribución incrementa el beneficio promedio, *sin contar los costes para ellos*.[55] En algunos de los ejemplos que he dado antes, la cooperación tenía efectos negativos sólo porque los costes para el cooperador eran más altos que la suma total de los beneficios. El utilitario desinteresado no se echará atrás aunque su acción tenga efectos negativos en este sentido. Sin embargo, se abstendrá de cooperar si su acción implica un perjuicio directo para otros. Muchos aspirantes a fundadores de movimientos sociales se han visto en ese aprieto. Como se ha señalado, tal vez teman impulsar a empleadores o autoridades a tomar medidas represivas que perjudiquen a otros miembros del grupo y no sólo a ellos mismos.

De no experimentar este problema, los utilitaristas desinteresados quizá sean lo bastante numerosos para llevar el número de cooperadores a OB, donde entran los utilitaristas plenos. Si los primeros son demasiado escasos o el aprieto que he mencionado los disuade de actuar, se requieren *cooperadores incondicionales* para generar el número crítico OB que incitará a los utilitaristas plenos. Esos actores adoptan una serie de forma: pueden ser kantianos, santos, héroes o fanáticos, o estar ligeramente tocados. Lo que tienen en común es que no actúan ni en función de las consecuencias esperadas de su acción, ni de la cantidad de los otros cooperadores. Los nazis fanáticos, por ejemplo, se incorporaron al NSDAP en un momento en que sus perspectivas eran poco más que castillos en el aire. Los utilitaristas, en contraste, actúan en función de las consecuencias esperadas de su acción. Aunque en sí misma su expectativa pueda ser una función del número de los demás cooperadores, este número no es el motivo directo de su acción (Capítulo 5). Si la curva OGCK de la Figura 24.3 no se extendiera bastante por debajo de O, ambos tipos de utilitaristas actuarían como precursores.

Con todo, hay una categoría adicional de actores que nunca desempeñarían ese papel. Se trata de los individuos cuya motivación es suscitada por la observación de la cooperación de otros o el conocimiento de que los cooperadores pueden observarlos (Capítulo 5). Un primer subconjunto de este grupo está motivado por la norma cuasimoral de la equidad: no es equitativo que nosotros permanezcamos a un costado mientras otros corren riesgos por nuestra causa común. La fuerza de esta motivación puede variar, en el sentido de que son distintas las cantidades de cooperadores que pueden suscitar la cooperación de diferentes individuos.

Aquellos cuya cooperación se pone en marcha debido a un umbral bajo, pueden por ello contribuir a estimular la de quienes tienen un umbral más alto. La motivación de un segundo subconjunto está constituida por las normas sociales. Si los no cooperadores pueden ser identificados y castigados con el ostracismo social, como suele suceder en los ámbitos laborales, por ejemplo, los cooperadores tal vez los induzcan a participar por vergüenza. La vergüenza también tiene grados. Algunos, avergonzados, quizá se decidan a sumarse a causa de una sola observación o acto de evitación, mientras que otros lo harán únicamente si quedan expuestos a la presión de muchos miembros del grupo, incluidos (tal vez especialmente) aquellos a quienes la vergüenza ha impulsado a cooperar.

Una categoría final la forman quienes se unen al movimiento por sus «beneficios de proceso»: porque es divertido o personalmente atractivo en otros sentidos. Para algunas personas, marchar y cantar por las calles, o saborear cualquier entretenimiento que se ofrezca, puede ser deseable en sí mismo, con prescindencia de la causa. Otras pueden ver la incorporación al movimiento como una oportunidad de «elevar la conciencia» o «formar el carácter». Aunque los estados deseados son en esencia subproductos (Capítulo 4), con lo cual se da a entender que quienes los buscan directamente pueden terminar por decepcionarse, es posible, no obstante, que las personas en cuestión se sumen por esa razón. Como estos individuos son exteriormente indistinguibles de los pertenecientes a las otras categorías, también ellos pueden servir para inducir la adhesión de quienes tienen umbrales elevados para participar.

Esta tipología de motivos no brinda una explicación de la acción colectiva exitosa; se limita a proponer un marco a esa explicación. Hay muchas mezclas motivacionales diferentes que pueden generar el efecto de bola de nieve requerido. En algunos casos, el movimiento tal vez no pueda reclutar suficientes cooperadores, si los precursores son demasiado pocos para atraer a los utilitaristas, o si unos y otros, en conjunto, son demasiado escasos para atraer a aquellos cuya cooperación está condicionada a la del resto. En otros casos, quizás el proceso nunca se ponga en marcha, porque no hay precursores. El resultado depende de la distribución de las motivaciones en la población, así como de la tecnología de la acción colectiva. También depende de la organización y el liderazgo, en aspectos que he pasado por alto aquí con el fin de concentrarme en el caso puro de las elecciones descentralizadas.

* * *

NOTA BIBLIOGRÁFICA

El experimento de los bienes públicos y la desintegración se describe en Ernst Fehr y Urs Fischbacher, «The nature of human altruism», *Nature*, 425, 2003, págs. 785-791. Sobre el efecto de bola de nieve en Europa oriental, véase Roger Dale Petersen, *Resistance and Rebellion: Lessons from Eastern Europe*, Cambridge y Nueva York, Cambridge University Press, 2001. Esta obra también muestra que varias motivaciones pueden existir juntas dentro de un mismo individuo, no sólo (como en mi texto) en diferentes individuos. Examino algunas variaciones de la tecnología en *The Cement of Society: A Study of Social Order*, Cambridge y Nueva York, Cambridge University Press, 1989 [*El cemento de la sociedad: las paradojas del orden social*, Barcelona, Gedisa, 1997]. Otras variaciones se analizan en Gerald Marwell y Pamela Oliver, *The Critical Mass in Collective Action: A Micro-Social Theory*, Cambridge y Nueva York, Cambridge University Press, 1993. Todd Sandler, *Collective Action: Theory and Applications*, Ann Arbor, University of Michigan Press, 1992, ofrece un elegante panorama general. Se encontrará una aplicación que toma en cuenta las diferencias que pueden existir entre los agentes tanto en recursos como en motivaciones en John R. Bowman, *Capitalist Collective Action: Competition, Cooperation, and Conflict in the Coal Industry*, Cambridge y Nueva York, Cambridge University Press, 1993. La cooperación en los cárteles sobre la base de la estrategia del gatillo se estudia en Charles Lipson, «Bankers' dilemmas: private cooperation in rescheduling sovereign debts», en Kenneth A. Oye (comp.), *Cooperation Under Anarchy*, Princeton, Princeton University Press, 1986, págs. 200-225. El caso del comercio de diamantes en Nueva York, examinado en el Capítulo 21, también es relevante aquí. Se hallará una descripción de la acción colectiva que hace hincapié en las relaciones verticales entre líderes y seguidores en Stathis N. Kalyvas, *The Logic of Violence in Civil War*, Cambridge y Nueva York, Cambridge University Press, 2006.

Notas al pie

[49] Garrison Keillor, *Lake Wobegon Days*, Nueva York, Viking, 1985.

[50] Dios podía verla, desde luego. En algunas iglesias medievales, muchos capiteles maravillosos están tan altos que nadie puede observar los detalles sin binoculares, que no existían cuando aquellos se construyeron. Pero como Dios podía verlos, no importaba. Los habitantes de Lake Wobegon no parecen haber tenido el mismo fervor religioso.

[51] Expondré las cosas como si las elecciones individuales fueran binarias (los individuos o bien cooperan o bien no cooperan) y los resultados fuesen continuamente variables (un bien público, como el aire puro, puede suministrarse en mayor o menor medida). En realidad, los individuos tal vez difieran en *la cantidad* de su aporte, y no meramente en *si lo hacen o no*. No tomaré en cuenta esta complicación. Además, algunos bienes públicos son «desproporcionados» o discretos. Si los miembros de una comunidad cabildean para mantener abierta la escuela local, ésta cerrará o bien quedará abierta. Esta complicación puede eludirse si se interpreta el resultado como la *probabilidad* continuamente variable de la provisión del bien público.

[52] Aquí y en otros lugares, palabras como «primero», «medio» y «último» pueden referirse a los momentos en que se suman sucesivos cooperadores, como en la construcción de un movimiento revolucionario. Pero también pueden aludir a actos simultáneos de cooperación, como en el voto. Decir que los últimos votantes suman poco es decir que el beneficio creado en una situación en la cual todos votan es casi igual al beneficio creado cuando casi todos votan.

[53] A diferencia del DP de reiteración finita, este enunciado se refiere a una situación de duración indefinida. Que en un caso dado las estrategias del gatillo formen un equilibrio dependerá de la estructura de retribuciones cardinales y del índice de descuento temporal de los agentes. En principio, otras estrategias también podrían generar un equilibrio cooperativo, pero sólo la estrategia del gatillo tiene la cualidad necesaria de punto focal.

[54] Para evitar complicar el análisis, sólo considero las variaciones en el conjunto de los individuos.

[55] Varios proverbios captan esta asimetría: «La amistad significa olvidar lo que das y recordar lo que recibes» y «Nunca olvides un servicio que te hayan prestado, y olvida de inmediato el que tú has brindado».

Capítulo 25

Toma colectiva de decisiones

Con frecuencia, los integrantes de un grupo –desde la familia hasta la sociedad en su conjunto– necesitan regular asuntos de interés común por medio de decisiones que son vinculantes para todos ellos. Volvamos a considerar la cuestión de la regulación del consumo de agua en períodos de escasez (Capítulo 5). A veces, este problema de acción colectiva puede resolverse mediante decisiones descentralizadas, a través de una combinación de normas morales, cuasimorales y sociales. Sin embargo, a menudo el concejo municipal se ve en la obligación de limitar el suministro de agua o reducir el consumo mediante la prohibición de ciertos usos, como el riego del césped o el llenado de las piscinas. Cuando la acción colectiva falla, quizá se requiera una toma colectiva de decisiones.

Veamos otro ejemplo, la práctica del voto en elecciones nacionales. La decisión de votar o quedarse en casa es un clásico problema de acción colectiva. Como sabe que su voto no significará ninguna diferencia en el resultado, cada ciudadano tiene un interés personal que dicta la abstención. No obstante, si todo el mundo se abstuviera, o la concurrencia a las urnas cayera a niveles muy bajos, la democracia misma podría correr el peligro de ser reemplazada por una dictadura o una oligarquía, contra el interés de (casi) todos. En muchas democracias, la concurrencia electoral alcanza de hecho niveles respetables, del 50% al 80%, como resultado de las decisiones descentralizadas de los ciudadanos. Es posible que algunos se pregunten: «Pero ¿qué pasaría si todo el mundo se abstuviera?» Otros tal vez se digan: «Como la mayoría se molesta en votar, es justo que yo

también lo haga». Y otros más acaso calculen que, «si bien la influencia de mi voto sobre la viabilidad de la democracia es minúscula, es importante si se la multiplica por la gran cantidad de ciudadanos que afecta». En una pequeña aldea, algunos pueden temer que, «si me quedo en casa, mis vecinos se darán cuenta y lo desaprobarán».

Si estas motivaciones, por separado y combinadas, demuestran ser demasiado débiles, la concurrencia electoral puede caer a niveles desastrosos, en un proceso que en parte se autorrefuerza («Como son muy pocos los que se molestan en votar, ¿por qué debería molestarme yo?»). Para invertir ese proceso, el Parlamento quizá disponga la obligatoriedad del voto e imponga una multa a quienes no concurren a votar, además de someter la ley a la aprobación de los votantes en un referéndum. Cuando votan para decidir si el voto debe ser obligatorio, los ciudadanos enfrentan una opción que es muy diferente de la que afrontan al contemplar la posibilidad de votar o no en elecciones ordinarias sin obligatoriedad del sufragio. Las opciones no son «voto» contra «me quedo en casa», sino «todo el mundo vota» contra «todo el mundo tiene la libertad de quedarse en casa».[56] Si muchos de los que prefieren la segunda opción en la primera alternativa prefieren la primera en la segunda, *decidirán en forma colectiva* que el voto sea obligatorio.

La toma colectiva de decisiones tiene que ver con la *elección de un curso de acción*. Antes de ingresar en el proceso de esa toma colectiva de decisiones, cada miembro tiene preferencias por tal o cual curso, que derivan de sus preferencias fundamentales, junto con una serie de creencias causales acerca de las relaciones entre medios y fines. El objetivo básico de la toma colectiva de decisiones es *agregar* las preferencias individuales en materia de planes de acción, mediante uno de los tres mecanismos que pronto examinaremos.[57] La agregación también puede inducir una *transformación* de dichas preferencias individuales, y generar un incentivo para que los individuos *distorsionen* las suyas propias. La interacción entre agregación, transformación y distorsión de las preferencias puede contribuir a crear una considerable complejidad.

En muchos de los casos que examinaré, un grupo más pequeño de individuos toma decisiones que son vinculantes para un grupo más grande. En ocasiones, se les ha delegado esa tarea, en cuanto representantes o negociadores del grupo mayor. De ser así, el conocimiento de que sus decisiones tendrán que ser ratificadas por sus representados, o de que no

conseguirán la reelección si no obtienen resultados satisfactorios, tal vez actúe como una coacción sobre ellos. En otros casos, la sociedad en general no tiene, salvo por medio de una revolución, poder para influir en quienes toman las decisiones que moldean su vida. No obstante, aun en estas circunstancias podemos hablar de toma colectiva de decisiones dentro de la elite. Luego de la muerte de Stalin, hubo un liderazgo colectivo del Politburó. Y la junta chilena que ejerció el poder entre 1973 y 1990 tenía un mecanismo interno sumamente estructurado de toma colectiva de decisiones.

Los tres mecanismos de agregación que revisaré son la *discusión*, la *negociación* y la *votación*. Creo que la lista es exhaustiva. Si bien algunas decisiones colectivas se toman mediante un dispositivo aleatorio, como cuando una asamblea lanza una moneda para romper un empate o los dirigentes se eligen por sorteo dentro de la población en general, la decisión misma de adoptar este procedimiento debe hacerse a través de la discusión, la negociación o el voto. No abordaré la enigmática cuestión de si uno de estos tres procedimientos es, en algún sentido, más básico que los otros: por ejemplo, si la decisión de resolver por votación una cuestión determinada debe tomarse tras una discusión. Investigaré, en cambio, las propiedades de cada procedimiento, incluidas sus patologías importantes.

Querría ante todo dar algunos ejemplos de los tres procedimientos. La discusión pura se observa (o se presume al menos que es la regla) en los jurados a los que se exige una decisión unánime. Aun en este caso, algunos de sus miembros tal vez recurran a una negociación tácita en virtud de su mayor aptitud para persistir, esto es, su menor impaciencia para terminar con la labor del jurado y volver a su vida cotidiana.[58] Como el tiempo siempre cuenta cuando es preciso tomarse una decisión, y los participantes en el proceso a menudo descuentan el futuro a velocidades diferentes, es posible que, en realidad, este caso sea típico.

La negociación pura tiene su ilustración en los juegos secuenciales de «reparto de un dólar», en los que las partes hacen ofertas y contraofertas sucesivas. El resultado está determinado por el mecanismo y la capacidad de negociación de las partes, es decir, los recursos que les permiten hacer amenazas y promesas creíbles. La Figura 20.2 representa el proceso.

La votación pura era la concepción que Rousseau se hacía de la toma colectiva de decisiones. Los ciudadanos debían formarse sus preferencias aislados unos de otros, a fin de que la elocuencia y la demagogia no los

contaminaran. Como también votarían de manera aislada, el tráfico de votos quedaba excluido. En los sistemas políticos reales, este ideal nunca se alcanza. Es posible ilustrarlo, tal vez, con ciertas decisiones de bajo coste, como la elección de miembros a una academia científica, cuya principal función es elegir a los nuevos miembros.

La mezcla de discusión y votación, sin negociación, puede ilustrarse con las decisiones en materia de contrataciones y ejercicio de cargos en un departamento universitario. Se supone que dichas decisiones están exclusivamente gobernadas por la deliberación sobre los méritos del candidato, seguida de una votación. Aunque este ideal no siempre corresponde a la realidad, a veces se alcanza. En los buenos departamentos hay una norma contra el intercambio de favores, reforzada por otra contra los votos que no se fundamentan.

La mezcla de discusión y negociación, sin votación, tiene su ejemplo en las negociaciones salariales colectivas. Cuando un sindicato y la patronal deciden cómo dividir los ingresos de la empresa, tal vez parezca que sólo hay una negociación. Sin embargo, una observación más detallada revelará que siempre hay una discusión sustancial sobre cuestiones fácticas, como la solidez financiera de la empresa y la productividad de la mano de obra.

La mezcla de negociación y votación se da cuando los miembros de un sindicato tienen que ratificar, mediante el voto, un acuerdo salarial negociado por sus representantes. En tales situaciones, el resultado esperado de la votación actúa como el aspecto amenazante de la negociación.

La toma de decisiones políticas, sea por parte de una comisión, una asamblea o la población en general, a menudo abarca los tres procedimientos.[59] Una vez más, este hecho se deduce de la necesidad de llegar a una decisión lo antes posible. La votación suele plantearse cuando un problema debe decidirse en forma urgente, razón por la cual los participantes no tienen tiempo de deliberar hasta llegar a una decisión unánime. En términos más prosaicos, quizá no se sientan motivados a buscar la unanimidad. Si la decisión es más urgente para algunos participantes que para otros, también se suscita la posibilidad de negociar, dado que quienes son más capaces de darse el lujo de esperar pueden exigir concesiones a cambio de una pronta decisión. En las comisiones y asambleas permanentes, la negociación también surge a través del intercambio de favores, debido a la desigual intensidad de las preferencias en torno de los problemas que deben conciliarse. Entre otros mecanismos

de negociación parlamentaria pueden mencionarse el obstruccionismo y la «política de la banca vacía», mediante la cual un grupo puede explotar las reglas del quórum para obtener lo que no podría conseguir por otros medios.

En tales casos, las fuentes del poder de negociación se crean dentro de la misma asamblea. En otros, los decisores pueden basarse en recursos que existen con independencia de la asamblea: dinero y personal. En 1789, los debates de la Asamblea Constituyente francesa estaban suspendidos entre las tropas del rey y las multitudes parisinas; en un comienzo, estas últimas eran un arma que los diputados podían esgrimir contra las primeras, y luego fueron una amenaza que algunos de los representantes desplegaban contra otros. En 1989, las conversaciones cuasiconstitucionales o preconstitucionales de la Mesa Redonda en Polonia estaban suspendidas entre la amenaza de intervención soviética y la perspectiva de una parálisis económica. Si un voto no puede comprarse con la promesa de otro voto, como en el intercambio de favores, es posible hacerlo con dinero, mediante la asignación, por ejemplo, de fondos partidarios con destino a una campaña por la reelección.

Como este examen permitirá advertir con claridad, los tres modos de toma colectiva de decisiones pueden verse como tres pasos en una secuencia idealizada, en el sentido de que cada uno de ellos surge naturalmente del precedente. Aunque la discusión tiene como objetivo intrínseco la unanimidad, en cuanto se basa en razones que se suponen válidas para todos, este fin se alcanza en contadas ocasiones. Para zanjar la cuestión, es preciso votar. Y como la votación suele tener lugar entre individuos que deben decidir sobre muchos temas, da naturalmente origen a la negociación bajo la forma del intercambio de favores.

Discusión

La discusión es el esfuerzo por persuadir mediante razones. Desde el elogio de Atenas hecho por Pericles, este modo de toma de decisiones ha guardado íntima relación con la política democrática:

> Nuestros hombres públicos tienen que atender, además de la política, sus asuntos privados, y nuestros ciudadanos corrientes, aunque ocupados en los afanes de su industria, son no obstante jueces justos de los asuntos públicos; pues, a diferencia de cualquier otra nación, nosotros consideramos al ciudadano que no participa de esos deberes no como falto de ambición sino

> como inútil, y somos capaces de juzgar las propuestas aun cuando no podamos presentarlas; en vez de ver la discusión como un obstáculo en el camino de la acción, la estimamos como un preliminar indispensable para cualquier acción prudente.

El vínculo entre la institución del debate público y la «acción prudente» puede ser un tanto indirecto. Con frecuencia, el principal efecto del marco público es la exclusión de las apelaciones abiertas al interés. En un debate público, un orador que dijera: «Debemos hacer esto porque es bueno para mí» no convencería a nadie; más aún, quedaría sujeto a unas sanciones informales y un ostracismo que le quitarían eficacia en el futuro. El marco público obliga aun a quienes están exclusivamente motivados por el interés a presentar sus propuestas políticas como si las motivaran valores más imparciales. Este proceso de *distorsión* de las preferencias difiere del de *trasmutación* (Capítulo 4) del mismo modo como el engaño difiere del autoengaño. Es el interés del orador, no su necesidad de autoestima, el que lo lleva a distorsionar ese mismo interés y presentarlo como razón. Es posible que su interés también lo lleve a hacer que la distorsión sea difícil de advertir, al propiciar en términos imparciales un curso de acción que se desvía un tanto (pero no demasiado) del que coincidiría perfectamente con dicho interés. En realidad, si fuera demasiado obvia, la distorsión podría resultar contraproducente. Por ello, *la necesidad de disimular nuestra preferencia fundamental puede provocar un cambio en nuestra preferencia en materia de cursos de acción*, por obra de lo que podemos concebir como la «fuerza civilizadora de la hipocresía».

En muchas sociedades, la propiedad se ha utilizado como un criterio para el sufragio. Es posible, desde luego, proponer argumentos imparciales en favor de ese principio. En la Convención Federal, Madison sostuvo que las rigurosas limitaciones impuestas por el Senado en lo concerniente a la propiedad, en vez de proteger a los privilegiados contra el pueblo, eran un recurso para proteger al pueblo contra sí mismo. Pero, como se ha señalado, hay algo intrínsecamente sospechoso en estos argumentos, que coinciden demasiado bien con el egoísmo de los ricos. Tal vez sea útil entonces referirse al alfabetismo como un criterio imparcial que, *aunque de manera imperfecta, está muy* correlacionado con la propiedad. En varios momentos de la historia de los Estados Unidos, el alfabetismo también ha actuado como sucedáneo legitimador de otras metas inconfesables, como el deseo de mantener a los negros o los católicos al margen de la política. La política inmigratoria estadounidense también lo ha utilizado como sustituto de

criterios que no podían enunciarse públicamente. Las propuestas de seleccionar a los inmigrantes mediante pruebas para determinar si sabían leer y escribir en su lengua natal solían justificarse como una manera de elegir sobre la base del mérito individual, un procedimiento imparcial de amplia aceptación. La verdadera motivación de los defensores del alfabetismo era, sin embargo, el prejuicio o el interés grupal. Los nativistas patricios querían excluir a los inmigrantes habitualmente analfabetos de Europa central y sudoriental. Y la clase obrera temía que una afluencia de trabajadores no calificados provocara una reducción de los salarios.

Sería sin duda erróneo suponer que la discusión *siempre* puede reducirse a una manera más o menos sutil de promover el interés propio. Si así fuera, la distorsión no tendría sentido porque nadie se la tragaría jamás.[60] Si los oradores están motivados por un deseo sincero de promover el bien público, la argumentación y el debate pueden modificar sus creencias de un modo que induce un cambio en las preferencias de cursos de acción. Es especialmente probable que esto ocurra si los diversos integrantes de un grupo tienen acceso a informaciones diferentes, de modo que, para mejorar la calidad de sus decisiones, pueden hacer un fondo común con su conocimiento.[61] Si el grupo integra un cuerpo representativo, es importante seleccionar delegados con antecedentes muy diferentes. En la elección de representantes para una asamblea nacional, por ejemplo, esta consideración habla en favor de la votación proporcional con un umbral bajo o sin ningún umbral.[62] También se podría exigir que los representantes de un distrito electoral fueran residentes de esa comunidad.

Aunque tal vez con menos asiduidad, la gente discute las preferencias fundamentales y las modifica como resultado del debate. A menudo, los cambios se producen gracias al descubrimiento de similitudes ocultas entre casos o la revelación de similitudes superficiales. Muchas personas se oponen, por ejemplo, al uso obligatorio de «órganos cadavéricos» para los trasplantes. Creen que si la familia tiene objeciones religiosas a este procedimiento, sería menester respetar sus sentimientos. Contra este punto de vista, podría aducirse la realización obligatoria de una autopsia en caso de muerte sospechosa, aun cuando el procedimiento sea contrario a las creencias religiosas de la familia. Si lo indicado para determinar la causa de una muerte es tomar medidas invasivas, podría argumentarse que éstas también deben ser aceptables cuando el objetivo es salvar una vida. El cambio puede producirse asimismo cuando se estima que un principio

general contradice las intuiciones referidas a casos particulares. Una persona quizás aceptara el uso obligatorio de órganos cadavéricos por razones utilitarias, pero rechazaría la implicación de que es lícito matar a un individuo escogido al azar y utilizar su corazón, sus riñones, sus pulmones y su hígado para salvar la vida de otros cinco.[63] Como resultado, un utilitarismo inicial irrestricto podría revisarse para tomar en cuenta los valores no consecuencialistas (Capítulo 4).

Sin embargo, los efectos de *hablar ante un auditorio* pueden debilitar los beneficios de la discusión. Los individuos orientados a la actuación pública pueden, no menos que otros, estar sometidos a un amor propio que los hace renuentes a admitir en público que han cambiado de opinión. En el Capítulo 3 señalé que ésa fue la principal razón aducida por Madison, mucho tiempo después, para explicar por qué la Convención Federal había sesionado a puertas cerradas y se había impuesto la obligación de silencio a los delegados. Con todo, su argumento podría parecer estar en conflicto con un argumento tradicional que promueve la apertura al público de los debates de las asambleas. Muchas decisiones legislativas tienen un fuerte impacto de corto plazo sobre los intereses de los legisladores. Si el proceso de toma de decisiones se desarrolla a cubierto de la vista del público, la discusión sobre el interés común degenerará con facilidad en una abierta negociación de intereses. La autorización otorgada al público para seguir las sesiones y observar las votaciones tiende a limitar esas maniobras egocéntricas y, como subproducto, a promover el bien común. Según escribió Bentham: «Cuanto mayor es la cantidad de tentaciones a las que está expuesto el ejercicio del poder político, más necesario es dar a quienes lo tienen a su cargo las más vigorosas razones para resistirlas. Pero no hay razón más constante ni más universal que la vigilancia del público». O, como dijo Louis Brandeis, un juez estadounidense: «La luz del día es el mejor desinfectante».

Estas observaciones apuntan a señalar una tensión en el proceso de discusión. Si los debates se celebran en público, la calidad de la argumentación sufrirá un menoscabo. Si se realizan a puertas cerradas, la discusión puede degenerar en negociación. Sin embargo, la tensión puede mitigarse si las cuestiones por decidir dan poca cabida al juego de los intereses privados. Las asambleas constituyentes tal vez sean menos propensas a tomar decisiones en beneficio propio que las legislaturas ordinarias, no porque los delegados tengan motivaciones más imparciales,

sino porque (o en la medida en que) sus intereses tienen menos influencia sobre los asuntos en cuestión. Además, como ya he indicado varias veces, debido a la perspectiva de largo plazo implicada por la elaboración constitucional, es posible que el egoísmo remede hasta cierto punto el deseo de promover el bien público.

Votación

Cuando la discusión no logra generar consenso con respecto a un curso de acción, tal vez sea necesario votar. Los sistemas de votación varían enormemente. En la votación popular, las dimensiones de la variación incluyen el derecho al sufragio, el modo de votar (secreto contra abierto), la mayoría necesaria para alcanzar una decisión y, en algunos sistemas de referéndum, el quórum. En las votaciones en asambleas, las principales dimensiones son el quórum, la magnitud de la mayoría y la elección entre la votación por lista o a mano alzada (y procedimientos similares, como el voto «por aclamación» o «parados contra sentados»). La votación secreta en las asambleas es poco habitual, pero no inaudita. El presidente de la Asamblea Constituyente francesa de 1789-1791 fue elegido de ese modo. En épocas recientes, Italia y Colombia han puesto en práctica la votación secreta en el parlamento. Adviértase que ese tipo de votación debe distinguirse de las sesiones secretas en las que no se admite la presencia de espectadores. Estas últimas pueden combinarse con la votación pública que permite a los asambleístas hacer promesas creíbles de intercambio de favores, lo cual sería imposible con la votación secreta. En contraste, si las sesiones son abiertas al público, algunos espectadores quizá tengan una reacción negativa al ver que sus representantes votan contra su preferencia en una cuestión, dado que no pueden observar las ganancias que así resultan posibles en otra.

En lo que sigue me limitaré a la votación por mayoría. Aunque esta práctica no sea universal, la decisión misma de aprobar propuestas por una mayoría más grande, como los tres quintos o los dos tercios, tendría que tomarse, al parecer, por mayoría simple. Las asambleas constituyentes, que a menudo imponen mayorías calificadas para las futuras enmiendas constitucionales, utilizan en forma casi invariable la votación por mayoría simple en sus sesiones.[64] El modelo idealizado en el cual las asambleas constituyentes deciden por unanimidad, detrás de un velo de ignorancia,

que decidirán por votación mayoritaria una vez que se levante dicho velo, tiene escasa relevancia para la elaboración constitucional real. Pasaré por alto la cuestión del quórum, salvo para señalar que la abstención o la «política de la banca vacía» pueden ser utilizadas por una minoría para obstaculizar una decisión que se habría aprobado de haber sido tratada y votada.

La falta de consenso que hace necesario decidir por votación puede deberse a la divergencia de preferencias fundamentales o de creencias, o de ambas. Como ejemplo, considérense los debates sobre el unicameralismo y el bicameralismo en la Asamblea Constituyente francesa de 1789. En términos muy generales, ese cuerpo estaba formado por tres grupos de dimensiones más o menos iguales. La derecha reaccionaria quería atrasar el reloj para volver a la monarquía absoluta, el centro moderado aspiraba a una monarquía constitucional con fuertes controles sobre el parlamento, y la izquierda también quería una monarquía constitucional, pero con débiles controles sobre ese cuerpo. En lo concerniente al problema del bicameralismo, las constelaciones eran como las muestra, muy simplificadas, el Cuadro 25.1.

En definitiva, el bicameralismo fue derrotado por la alianza de reaccionarios y radicales. Este fenómeno general –el acuerdo político basado en diferencias de preferencias y diferencias de creencias que se anulan entre sí– es bastante común. Sería incluso posible llegar a la unanimidad sobre esa base,[65] aunque de tipo diferente de la que podría surgir en la «situación discursiva ideal» en la cual los oradores sólo están motivados por el bien común y se muestran dispuestos a escuchar los argumentos.

	Preferencias fundamentales	*Creencias*	*Preferencias políticas*
Reaccionarios	Desestabilizar el régimen	El bicameralismo estabilizará el régimen	Unicameralismo
Moderados	Estabilizar el régimen	El bicameralismo estabilizará el régimen	Bicameralismo
Radicales	Estabilizar el régimen	El bicameralismo desestabilizará el régimen	Unicameralismo

CUADRO 25.1

En mi versión estilizada del debate, una mayoría creía que el bicameralismo estabilizaría el régimen y una mayoría (diferente) quería estabilizarlo (véase el Cuadro 25.1). Si las decisiones colectivas se hubieran tomado, primero, mediante la agregación de las creencias a través de una votación (sincera) por mayoría, segundo, mediante la agregación de las preferencias fundamentales a través de una votación (sincera) por mayoría, y por último, mediante la acción que, de acuerdo con la creencia agregada, mejor realizara la preferencia agregada, se habría elegido el *bicameralismo*. (Para impedir este resultado, los reaccionarios podrían haber hecho pública la falsa creencia de que el bicameralismo desestabilizaría el régimen, con lo cual habrían creado una mayoría para esa creencia y, por ende, una mayoría para la elección del unicameralismo.) Que yo sepa, este procedimiento de «doble agregación» nunca se utiliza en la práctica. Dadas las complejidades de agregar directamente las preferencias políticas, es probable que cualquier sistema de ese tipo fuera imposible de manejar.

En el debate sobre el bicameralismo, los votantes diferían tanto en sus creencias como en sus objetivos finales. En otros casos puede coincidir en uno de estos dos aspectos y diferir en otro. Como lo que se observa y se agrega realmente son las preferencias políticas, tal vez sea difícil discriminar los dos factores que intervienen en su factura. En abstracto, es posible que tratemos, no obstante, de determinar los efectos de la votación

por mayoría en la agregación de las creencias (suponiendo objetivos idénticos) y de las preferencias fundamentales (suponiendo creencias idénticas). Según Tocqueville, la democracia (es decir la votación por mayoría con un derecho amplio al sufragio) es superior a los otros sistemas en ambos aspectos. «El imperio moral de la mayoría se funda en parte sobre la idea de que hay más luz y cordura en muchos hombres reunidos que en uno solo, en el número de los legisladores que en su selección […] [y] en el principio de que los intereses del mayor número deben ser preferidos a los del menor.» Los funcionarios democráticos pueden «cometer […] graves errores», pero «no seguirán jamás sistemáticamente una tendencia hostil a esa mayoría».[66]

Consideremos en primer lugar la agregación de creencias. Desde hace mucho se debate si el mejor procedimiento para llegar a creencias correctas es un derecho amplio o restringido al sufragio: si los muchos son más sabios que los pocos. Según Aristóteles, la cuestión pasaba por ponderar la cantidad (el número de participantes en el proceso político) contra la calidad (la competencia de los participantes):

> La calidad puede existir en una de las clases que constituyen el Estado y la cantidad en otra. Por ejemplo, los malnacidos pueden ser más numerosos que los nobles, o los pobres más que los ricos, no obstante lo cual la clase más abundante tal vez no sea tan superior en cantidad como inferior es en calidad. Por ende, es preciso juzgar estos dos factores por comparación entre uno y otro. Donde, en consecuencia, la multitud de los pobres es superior en la proporción expuesta [a fin de compensar su calidad inferior], es natural que haya democracia.

En lenguaje moderno, la cuestión puede enunciarse en términos del «teorema del jurado» de Condorcet. Supongamos que los miembros de un jurado exponen sus creencias (independientes) sobre si el acusado ha hecho lo que el fiscal afirma que ha hecho, y cada uno de ellos tiene una posibilidad de más del 50% de estar en lo cierto. Condorcet mostró que si el jurado decide mediante una votación por mayoría, su posibilidad de acertar aumenta cuanto más numeroso es ese cuerpo,[67] y converge en la certeza cuando éste es indefinidamente grande. Además, para un tamaño dado del jurado, la posibilidad de que la mayoría acierte aumenta cuando también lo hace la posibilidad de que cada uno de sus miembros dé en la tecla.[68] Por lo tanto, como sugería Aristóteles, se puede mejorar el resultado sea aumentando el número de jurados, sea elevando sus calificaciones.[69]

Si vamos más allá de Aristóteles, podemos observar que las calificaciones pueden ser una función directa del número y no de la

posición socioeconómica. En el lenguaje de las ciencias sociales, la competencia de los votantes tal vez sea «endógena» al sistema en vez de darse «exógenamente». Supongamos que tenemos que elegir entre dos sistemas políticos, la oligarquía y la democracia, y que en ambas la decisión se toma por votación mayoritaria pero con diferentes cantidades de personas habilitadas para votar. En una democracia, los votantes decidirán racionalmente mantenerse ignorantes, dado que la influencia de cada uno sobre el resultado es muy pequeña.[70] En una oligarquía, los votantes invertirán más en la búsqueda de informaciones, dado que cada uno de ellos tiene mayor influencia.

Bentham señaló que este argumento también es valedero para las votaciones en una asamblea: «Cuanto mayor es el número de votantes, menores son el peso y el valor de cada voto, más pequeño su precio a ojos del votante, y menor el incentivo que éste tiene para asegurarse de que se ajusta al verdadero fin, e incluso para emitirlo». En respuesta al argumento de que una asamblea (evocaba la Asamblea Constituyente francesa de 1789) debía ser numerosa, dado que «la probabilidad de prudencia en el juicio aumenta con la cantidad de miembros», escribió que «la reducción que esta misma causa provoca en la fuerza de la motivación para actuar de manera esclarecida contrarresta esa ventaja». En este balance entre la cantidad y la calidad habrá un tamaño óptimo del electorado que maximice la probabilidad de que la votación por mayoría produzca la creencia correcta.[71]

Consideremos a continuación la agregación de preferencias mediante la votación. Los dos principales problemas que se plantean, en estrecha relación, son la *distorsión* de las preferencias y la *indeterminación* del resultado. Empezando por el primero, es posible que la gente tenga un incentivo para votar propuestas o candidatos distintos de los que preferiría ver aprobados o elegidos. La adopción del sufragio abierto y no cerrado puede inducir este fenómeno. En la Atenas clásica, la mayoría de las decisiones de la asamblea se tomaban a mano alzada, y como consecuencia de ello es posible que algunos ciudadanos temieran votar de acuerdo con su verdadero parecer. Así, Tucídides afirma que, «con el entusiasmo de la mayoría [por la expedición siciliana], los pocos que no la apreciaban temían mostrarse poco patrióticos si levantaban su mano contra ella». (En contraste, las decisiones en los grandes tribunales atenienses con jurado se tomaban mediante sufragio secreto.) La adopción de un voto por lista y no

por otros métodos, como el de «parados contra sentados», también puede intimidar a los votantes. En las asambleas constituyentes de París (1789-1791) y Fráncfort (1848), los radicales demandaban habitualmente el voto por lista en los asuntos importantes, con la amenaza implícita y a veces explícita de que expondrían a quienes votaran contra las propuestas radicales a la violencia popular, haciendo circular listas con sus nombres. Aun cuando hubiera una clara mayoría con el sistema de «parados contra sentados», que hacía difícil identificar cómo votaba cada individuo, el resultado podía revertirse en una votación por lista.[72]

La distorsión también puede aparecer con el sufragio secreto. Prácticamente en todos los sistemas de votación pueden surgir situaciones en las cuales un votante, al inclinarse por una alternativa que no es su predilecta, es capaz de generar un resultado que prefiere al que se habría producido si hubiese votado con sinceridad.[73] (Podría plantearse una excepción si los candidatos o las propuestas fueran escogidos mediante un dispositivo aleatorio, donde la probabilidad de elección de una alternativa fuera igual a la proporción de votantes que la favorecen. En este caso, no surgiría el problema del «voto desperdiciado». Las desventajas del sistema son evidentes y explican por qué nunca se lo ha adoptado.) El deseo de ver a nuestra opción de primer orden ganadora por un margen no demasiado amplio puede inducirnos a votar contra ella. En el Capítulo 18 mencioné, por ejemplo, que los socialistas podrían votar por los comunistas con el objeto de inclinar la plataforma de su propio partido hacia la izquierda. Si existe la certeza de que nuestra opción de primer orden no será elegida, tal vez votemos por la mejor alternativa entre las que tienen posibilidades de ganar. Algunos sistemas de votación también generan un incentivo para situar a un candidato o propuesta preferidos por otros votantes en una posición menos favorable de lo que dictarían nuestras verdaderas preferencias (véase más adelante el ejemplo correspondiente), o para plantear nuevas alternativas con el propósito exclusivo de hacer más probable la elección de nuestra alternativa predilecta.

En algunos casos, el resultado de la votación por mayoría es indeterminado. Supongamos que en una asamblea municipal hay tres bloques que son más o menos del mismo tamaño y que representan, respectivamente, a la comunidad empresarial, a los trabajadores industriales y a los profesionales de los servicios sociales. La asamblea debe escoger entre construir una piscina techada, otorgar un subsidio a la orquesta

sinfónica local o construir una pista de golf. De conformidad con el estereotipo de estos grupos, supongamos que (luego de prolongados debates) ordenan las opciones como se muestra en el Cuadro 25.2.

Si las alternativas se cotejan unas con otras en votos pareados, hay una mayoría de empresarios y trabajadores que prefieren la pista de golf a la orquesta, una mayoría de empresarios y profesionales que prefieren la orquesta a la piscina y una mayoría de profesionales y trabajadores que prefieren la piscina a la pista de golf. Por consiguiente, las «preferencias sociales» son *intransitivas* o *cíclicas*. En el caso de la elección individual, la transitividad era una exigencia de la racionalidad (Capítulo 11). En el presente contexto, la cuestión no tiene que ver tanto con la racionalidad como con la determinancia. Si el concejo municipal sólo debe atenerse a las clasificaciones del Cuadro 25.2, cuesta ver cómo podría tomar alguna decisión. Como la votación se ha realizado porque el concejo era incapaz de llegar a un consenso, es improbable que la continuación del debate sirva de algo. Si pudiéramos medir la *intensidad* con que los diversos grupos prefieren una de las opciones a otra, o la medida en que éstas satisfacen necesidades objetivas, tal vez podríamos decir que una opción es inequívocamente superior a las demás. Sin embargo, no existe ningún procedimiento general que nos permita comparar grados de intensidad de preferencias o de satisfacción de necesidades entre los distintos individuos.[74] Es inútil preguntarles, por ejemplo, cuánto valoran las opciones en una escala de cero a diez. Por un lado, no podemos saber si un puntaje dado (siete, por ejemplo) significa lo mismo para los integrantes de los tres grupos. Por otro, al pedirles que ordenen las opciones les daríamos un incentivo para distorsionar la intensidad de sus preferencias, al asignar, pongamos por caso, diez a su opción predilecta y cero a las otras.

	Empresarios	*Trabajadores*	*Profesionales*
Pista de golf	1	2	3
Orquesta	2	3	1
Piscina	3	1	2

Cuadro 25.2

No está claro qué importancia tiene en la práctica este problema de las «preferencias sociales cíclicas». No puede plantearse si las preferencias individuales tienen un «único pico», esto es, si las opciones pueden clasificarse de la «más alta » a la «más baja» de tal manera que las preferencias de cada individuo se incrementen uniformemente al acercarnos a su curso de acción preferido y se reduzcan del mismo modo a medida que nos alejamos de éste. En muchos casos, hay una propiedad razonable de las preferencias. Si la escala preferida de gravámenes impositivos de un individuo es el 20%, preferirá el 19% al 18% y el 21% al 22%. Por otra parte, no hay ejemplos de una asamblea que simplemente baje los brazos y declare que, como no hay «voluntad popular», no se ha de tomar ninguna decisión. De hecho, si el statu quo es una de las opciones, esta idea es incoherente. Siempre se toma alguna decisión, sea por defecto (mantenimiento del statu quo), adopción de un procedimiento tradicional de votación o manipulación de la agenda.

Con todo, el hecho de que se llegue a una decisión no implica que ésta encarne en un sentido no arbitrario la voluntad popular o «general». Para una constelación de preferencias (sinceras) como la presentada en el Cuadro 25.2, la idea misma de voluntad general carece de significado. ¿Con cuánta frecuencia se producen esas constelaciones? Los científicos políticos han propuesto una serie de ejemplos. Otros han sostenido que la descripción de los presuntos ejemplos es errónea, y que un examen más detenido refuta estas ideas específicas sobre las mayorías cíclicas. Me referiré a dos casos que parecen ser auténticas instancias de preferencias cíclicas.

El 8 de octubre de 1992, el parlamento noruego decidió que el futuro aeropuerto del área metropolitana de Oslo debía construirse en Gardermoen (me referiré a esta opción como la alternativa G). Otras opciones eran Hobøl (alternativa H) y una solución que implicaba una combinación de Gardermoen y el aeropuerto de Fornebu, ya existente (alternativa D). Las opciones no iban a cotejarse entre sí, sino a considerarse de manera sucesiva con respecto al statu quo. Una vez que una de ellas recibió una mayoría de los votos, se la aprobó.[75] Aunque este voto serial era el sistema de votación tradicional en el parlamento, hay otros sistemas posibles; por ejemplo, cotejar entre sí las opciones en votos por pares hasta que surja una ganadora. Con la votación sucesiva, el orden en que se votan las opciones puede ser decisivo, como pronto veremos.

Las preferencias partidarias *expresadas*, que con excepciones poco importantes coincidieron con los votos de los diputados, eran las siguientes:

Partido Laborista (sesenta y tres diputados): G > D > H.
Una coalición del Partido de la Izquierda Socialista, la Democracia Cristiana y el Partido Agrario (cuarenta y dos diputados): D > H > G.
Partido Conservador (treinta y siete diputados): H > G > D.
Partido del Progreso (veintidós diputados): H > D > G.
Un diputado independiente: G > H > D.

Suponiendo que éstas eran preferencias *sinceras*, las preferencias sociales fueron cíclicas: D derrotaba a H por 105 a 60; H derrotaba a G por 101 a 64, y G derrotaba a D por 101 a 64. Antes de votar, el parlamento decidió, también por votación, el orden en que debían examinarse las alternativas. El laborismo propuso G-D-H, mientras que el presidente del parlamento sugirió D-H-G. Cuando estas propuestas se sometieron a votación, se impuso la del laborismo. Si hubiera vencido la del presidente del parlamento, los laboristas probablemente habrían votado por D, dado que de lo contrario su incapacidad para reunir una mayoría favorable a esta alternativa habría conducido a la adopción de la propuesta que menos les convencía, H. De acuerdo con el orden adoptado, el Partido Conservador se encontraba en un aprieto similar. En definitiva, los conservadores votaron por G, porque, de haber votado en contra de esta opción, se habría impuesto la alternativa que ponían en último lugar, D. Aunque en abstracto es posible que los laboristas fueran insinceros al situar D como su segunda opción, y que lo hicieran con el único fin de inducir a los conservadores a creer que el voto contra G provocaría la adopción de D, no hay pruebas en ese sentido. Si así fueran efectivamente las cosas, las preferencias sociales no serían cíclicas, pues H se impondría tanto a D como a G.

En el segundo ejemplo queda prácticamente excluido que las preferencias cíclicas sean un mero artefacto surgido de la distorsión. La cuestión se planteó en el contexto de la decisión sobre el orden de desmovilización del ejército estadounidense luego de la Segunda Guerra Mundial. La salida rápida era un bien escaso, que debía asignarse equitativamente. Para determinar los criterios, el ejército realizó encuestas en gran escala entre los hombres enrolados. En una encuesta en la cual los criterios se cotejaron en comparaciones por pares, las clasificaciones

mostraron cierta inconsistencia colectiva. Así, el 55% creía que un hombre casado con dos hijos que no había intervenido en combate debía ser dado de baja antes que un soltero con dos campañas en su haber; el 52% estimaba que dieciocho meses en ultramar eran más importantes que dos hijos, y el 60% calificó dos campañas como más valiosas que dieciocho meses de servicio en ultramar. Es sumamente improbable que los participantes distorsionaran sus preferencias.[76]

Negociación

La negociación es el proceso de concertación de un acuerdo por medio de amenazas y promesas creíbles. Un cónyuge puede amenazar con litigar por la custodia exclusiva de un hijo, si el otro padre no acepta la custodia compartida. En una negociación salarial, los trabajadores pueden amenazar con declarar la huelga, trabajar a reglamento o negarse a hacer horas extras, mientras que la patronal puede amenazar con interrumpir las tareas o cerrar las instalaciones. La dirección de una empresa tal vez amenace con despedir a un empleado a menos que éste se esfuerce más en su puesto. Un país quizás amenace con invadir a otro si éste no hace concesiones territoriales. En una asamblea constituyente, es posible que el delegado de una unidad territorial amenace con renunciar a menos que ese cuerpo sancione un modo de representación beneficioso para dicha unidad. Los senadores estadounidenses pueden amenazar con hacer obstruccionismo si el presidente no da marcha atrás con una designación. El congreso puede esgrimir la amenaza de negarse a votar el presupuesto si el presidente veta una ley.

En lo concerniente a las promesas, un miembro de un grupo que decide por votación puede prometer votar por una propuesta que es importante para uno de sus compañeros, con la condición de que éste vote por otra que es de importancia para él (intercambio de favores). El vendedor de una casa tal vez prometa no seguir ofreciéndola si un comprador acepta el precio que pide. De manera similar, es posible que un secuestrador prometa liberar a la víctima una vez que se pague el rescate, en vez de retenerla y hacer una nueva demanda. A la inversa, un gobierno puede prometer dejar salir de la cárcel a un terrorista una vez que los camaradas de éste hayan liberado a la víctima que tienen secuestrada. La víctima de un secuestro quizá prometa no describir la apariencia de los secuestradores a la policía a cambio de su

libertad. Una persona que se encuentra en una situación del dilema del prisionero puede prometer cooperar si la otra también lo hace.

El resultado de la negociación depende de la *credibilidad* de las amenazas y promesas. Una amenaza es creíble si es de esperar (por la razón que fuere) que el agente que la hace la lleve a cabo en caso de que la otra persona se niegue a cumplir. «La bolsa o la vida» es más creíble que «la bolsa o mi vida». Una promesa es creíble si es de esperar (por la razón que fuere) que el agente la cumpla una vez que la otra persona ha cumplido su parte. «Dibujaré las ilustraciones luego de que escribas el texto» es una promesa creíble, pues la persona que la hace tendrá un incentivo para mantenerla. En el juego de la confianza que examiné en el Capítulo 5 y en el Capítulo 15, el operador no tiene incentivos materiales para mantener la promesa de devolver la mitad de las ganancias generadas por la transferencia del inversor.

La credibilidad de las *amenazas* depende de factores tanto objetivos como subjetivos. Los factores objetivos pueden dividirse en opciones exteriores y opciones interiores. Una opción exterior es lo que una de las partes de una negociación obtendría si ésta se rompiera en forma irrevocable. En las negociaciones salariales, la opción exterior para los trabajadores sería el salario que podrían conseguir en otra empresa o el nivel de los beneficios por desempleo. En la negociación por la custodia de un hijo, las opciones exteriores se definen por el resultado esperado del litigio. En la Convención Federal de Filadelfia, la opción exterior consistía en lo que un estado podía obtener si se retiraba de la Unión (y concertaba tal vez una alianza con otra nación). La amenaza de abandonar una negociación sólo es creíble si quien la hace puede mostrar que estará mejor fuera de la relación de lo que lo estaría si aceptara la oferta de la otra parte.

Una opción interior es lo que las partes tienen a su disposición durante el proceso mismo de negociación, un factor determinante del tiempo que pueden mantenerse firmes. Para los trabajadores, la opción interior puede estar dada por la magnitud del fondo de huelga. Para los padres, podría ser la asignación de la custodia temporaria a uno de ellos. En Filadelfia, la opción interior estaba determinada por la avidez de los delegados por volver a los asuntos urgentes en sus estados natales. Las opciones interiores afectan la credibilidad de las amenazas porque determinan si una parte tiene incentivos para concretarlas. Una amenaza de huelga formulada por los trabajadores tal vez no sea creíble si (la empresa sabe que) éstos carecen de

fondo de huelga y en su mayor parte están casados y deben pagar elevadas cuotas hipotecarias.

En términos más generales, la amenaza de imponer un daño a otra parte no es creíble si, al hacerlo, uno también se inflige un perjuicio sustancial a sí mismo. Los miembros del congreso no pueden formular la amenaza creíble de negarse a votar el presupuesto si el presidente sabe que esta medida significaría un grave perjuicio tanto para su reputación como para su capacidad de acción. Lanjuinais, un delegado a la Asamblea Constituyente francesa de 1789, sostuvo que la futura legislatura no sería capaz de utilizar su control sobre los ingresos para impedir que el rey usara su poder de veto. Puesto que «suspender el pago de impuestos» con esa finalidad sería «como cortarse el cuello para curar una herida en la pierna», la amenaza de hacerlo no era creíble.

En el aspecto subjetivo, las amenazas son más creíbles si el agente tiene (y la otra parte cree que tiene) un prolongado horizonte temporal. Los negociadores más pacientes cuentan con una ventaja, dado que sus oponentes quizás estén dispuestos a negociar la *magnitud* de una concesión en función del *momento* en que se haga.[77] En las negociaciones entre los Estados Unidos y Vietnam del Norte celebradas en París, este último país jugó una buena apertura cuando su delegación alquiló una casa por *dos años*, como una manera de señalar que nada los apuraba. La neutralidad ante el riesgo o un bajo nivel de aversión a éste quizá den también a una de las partes cierta ventaja en las negociaciones. Un progenitor adverso al riesgo que tiene buenas posibilidades de obtener en los tribunales la custodia de su hijo tal vez acepte, no obstante, compartirla (Capítulo 11). Otro factor subjetivo que puede prestar credibilidad a las amenazas es el hecho de atribuir un valor relativamente bajo a lo que la otra parte tiene para ofrecer, en comparación con aquello a lo que se nos pide renunciar. En las negociaciones de divorcio, un padre quizá sea capaz de forzar un trato económico favorable porque el otro progenitor se preocupa mucho por conseguir la custodia del hijo.

El hecho de tener un extenso horizonte temporal también afecta la credibilidad en otro aspecto. Si no logro llevar a la práctica una amenaza en una interacción presente porque sería muy costoso hacerlo, mi aptitud de hacer amenazas creíbles en futuras interacciones se verá menoscabada. Tal vez sea más fácil entender la idea si se la expone a la inversa: puedo llevar a cabo una amenaza en una interacción presente aun cuando no tenga

incentivos para hacerlo, con el fin de forjarme la reputación de ser una persona cuyas amenazas deben tomarse en serio. Si la otra persona entiende la lógica de este argumento, quizá yo no tenga que concretar ninguna amenaza. No hay construcción de la reputación en un sentido literal; sólo la hay en un sentido «virtual». Más comúnmente, podría tener que plantear algunas amenazas «contra mis intereses» para señalar al mundo que soy ese tipo de persona. En los comentarios sobre el juego de la cadena de tiendas del Capítulo 20 me referí a ambas posibilidades.

La credibilidad de las *promesas* también depende de factores objetivos y subjetivos, aunque en aspectos un tanto diferentes. Un factor crucial para que una promesa sea creíble es la capacidad del agente de cumplirla. Consideremos como ejemplo un intento fallido de intercambio de favores en la Asamblea Constituyente francesa en el otoño de 1789. En tres reuniones entre el líder de los moderados, Mounier, y los radicales Barnave, Duport y Alexandre Lameth, estos últimos hicieron la siguiente propuesta. Ofrecerían a Mounier un poder de veto absoluto para el rey y el bicameralismo, si a cambio él aceptaba que el monarca renunciara a su derecho de disolver la asamblea, que la cámara alta tuviera únicamente un veto suspensivo y que hubiese convenciones periódicas para la revisión de la Constitución. Mounier se negó de plano, posiblemente porque no creía en la capacidad de sus tres interlocutores de cumplir su promesa, dado que la asamblea no tenía partidos en el sentido moderno de agrupamientos disciplinados a los que puede hacerse votar en bloque.

Otro ejemplo es el de las promesas de inmunidad procesal para las autoridades salientes en las transiciones a la democracia. Promesas de ese tipo se hicieron, se aceptaron y se rompieron en la Argentina en 1983, en Uruguay en 1984 y en Polonia y Hungría en 1989. (En los países latinoamericanos, las amenazas de golpe militar obligaron luego a cumplirlas.) En retrospectiva, los generales y dirigentes partidarios deberían haber entendido que esas promesas no eran creíbles, dado que los líderes entrantes que las negociaban no podían garantizar que la justicia y las legislaturas las respetaran. En Polonia, los negociadores de la oposición en las conversaciones de la Mesa Redonda, pertenecientes al ala izquierda de Solidaridad, sostuvieron que *pacta sunt servanda*: las promesas deben cumplirse. Cuando la derecha del movimiento subió al poder, ignoró el compromiso.

Entre las condiciones subjetivas de la credibilidad se cuenta, una vez más, un prolongado horizonte temporal. Si el agente tiene un bajo índice de descuento temporal y sabe que volverá a interactuar con la persona a quien ha hecho una promesa (o con otras personas que pueden observar su comportamiento), tiene un incentivo para respetarla. El intercambio de favores en las legislaturas, por ejemplo, puede mantenerse en virtud de esas previsiones. (En las asambleas constituyentes, que sólo se reúnen una vez, el incentivo para cumplir una promesa de reciprocidad debe ser más débil, y por lo tanto deben hacerse menos promesas de ese tipo.) Un agente quizá pueda también generar un desincentivo para quebrar su promesa. Si bien los secuestradores suelen no tener razones para creer en la promesa de su víctima de que no revelará su identidad a la policía, el secuestrado puede hacerla creíble si les entrega una información nociva y verificable sobre sí mismo que *ellos* puedan revelar en caso de ser arrestados.

Como las condiciones subjetivas o los estados mentales que dan forma al resultado de la negociación no pueden observarse directamente, los negociadores tienen interés en distorsionarlos, mediante un comportamiento verbal o no verbal. Que sepamos, los negociadores norvietnamitas firmaron un alquiler prolongado con el mero fin de dar la *impresión* de que eran pacientes. En el intercambio de favores, cada parte exagerará la importancia de aquello a lo que se le pide renunciar, para obligar a la otra a hacer una gran concesión. Cuando los trabajadores afirman atribuir gran importancia a las costosas medidas de seguridad en el lugar de trabajo, lo hacen tal vez como una estrategia para justificar un sustancioso aumento salarial al precio de dejarlas de lado. En muchos casos, es posible que los intentos de engañar sean demasiado transparentes para funcionar. Si en medio de un proceso de divorcio uno de los padres pretende mostrarse muy preocupado por conseguir la custodia de los hijos para lograr un acuerdo económico favorable, el otro quizá pueda documentar una falta uniforme de interés en ellos antes de que el matrimonio comenzara a hundirse, o la reciente aceptación de un trabajo que implica muchos viajes. Sin embargo, un progenitor perspicaz podría prever este problema y sentar las bases de la pretensión de desvelarse por los hijos antes de que el otro advirtiera el inminente derrumbe del matrimonio.

Como las partes involucradas en la discusión, los negociadores pueden tener un incentivo para distorsionar su interés y presentarlo como si se basara en un principio. Con todo, el razonamiento subyacente a la distorsión

es diferente. En la discusión, las partes quieren evitar el oprobio de basar sus propuestas en el interés despojado de todo adorno. En la negociación, no hay oprobio alguno que afecte la manifestación del interés. Se supone que las empresas y los trabajadores se preocupan por las ganancias y los salarios, no por el bien común. No obstante, la formulación de sus exigencias en términos de principio quizá brinde una ventaja estratégica a los negociadores. Éstos pueden sostener que, al echarse atrás con respecto a una pretensión basada en principios, hacen una concesión más grande que si el mero interés estuviera en juego, y esperan por lo tanto concesiones de la misma magnitud de la otra parte. Sin embargo, si ambas partes utilizan esta táctica, la negociación puede fracasar.

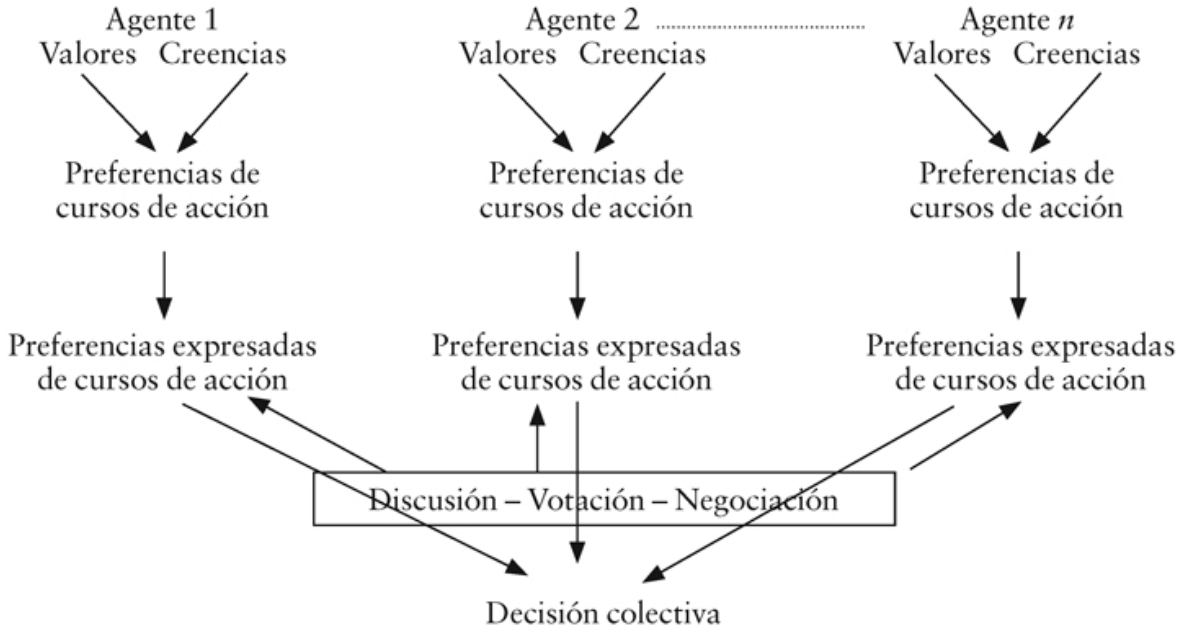


FIGURA 25.1

Sumario

Si atamos los diversos cabos de este capítulo, el proceso de toma colectiva de decisiones puede representarse como lo muestra la Figura 25.1. El punto central es tal vez que cada uno de los mecanismos de dicha toma colectiva, discusión, votación y negociación, crea un incentivo para distorsionar algún aspecto de nuestras preferencias. En otras palabras, un

mecanismo de agregación contribuye a dar forma a los insumos que el propio mecanismo ha de utilizar. Las preferencias expresadas en materia de cursos de acción son una función tanto de las preferencias reales de cursos de acción como del mecanismo que agrega aquéllas. En el aspecto del bienestar, el impacto de la distorsión es ambiguo. En virtud de la fuerza civilizadora de la hipocresía, los efectos quizá sean socialmente beneficiosos. En otros casos, el uso generalizado de esta táctica puede crear una situación del tipo del dilema del prisionero, en la cual todos pierden.

* * *

NOTA BIBLIOGRÁFICA

Con referencia a los mecanismos que en algunos aspectos son intermedios entre la acción colectiva y la toma colectiva de decisiones, véase Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge y Nueva York, Cambridge University Press, 1990 [*El gobierno de los bienes comunes: la evolución de las instituciones de acción colectiva*, México, Centro Regional de Investigaciones Multidisciplinarias, Universidad Nacional Autónoma de México/Fondo de Cultura Económica, 2000]. Se encontrará una investigación sobre los países con voto obligatorio y las sanciones que imponen a quienes no votan en http://www.idea.int/vt/compulsory_voting.cfm. Un esclarecedor aunque a veces excéntrico examen de la discusión y la votación es Jeremy Bentham, *Political Tactics*, Nueva York, Oxford University Press, 1999 [*Tácticas de las asambleas legislativas*, en *Obras selectas*, cuatro volúmenes, Buenos Aires, Librería «El Foro», 2003-2005, vol. 2]. Los pasajes citados de Bentham (traducidos del francés) pertenecen al igualmente interesante *Rights, Representation, and Reform*, Nueva York, Oxford University Press, 2002, págs. 35 y 122. Con respecto a la distorsión inducida por la deliberación, véase el Capítulo 5 de mi *Alchemies of the Mind: Rationality and the Emotions*, Cambridge y Nueva York, Cambridge University Press, 1999 [*Alquimias de la mente: la racionalidad y las emociones*, Barcelona, Paidós, 2002]. En lo concerniente a la diferencia entre las legislaturas ordinarias y las asambleas constituyentes, véase mi «Legislatures as constituent assemblies», en Richard Bauman y Tsvi Kahana (comps.), *The Least Examined Branch: The Role of Legislatures in the Constitutional*

State, Cambridge y Nueva York, Cambridge University Press, 2006, págs. 181-197. Acerca del voto sobre el bicameralismo en 1789, véase Jean Egret, *La Révolution des notables: Mounier et les monarchiens, 1789*, París, Armand Colin, 1950. El procedimiento de «doble agregación» se analiza en Christian List, «The discursive dilemma and public reason», *Ethics*, 116(2), enero de 2006, págs. 362-402. La discusión sobre el teorema del jurado de Condorcet se basa en Drora Karotkin y Jacob Paroush, «Optimum committee size: quality-versus-quantity dilemma», *Social Choice and Welfare*, 20(3), junio de 2003, págs. 429-441. Se encontrará un tratamiento completo del sufragio secreto en Hubertus Buchstein, *Öffentliche und geheime Stimmabgabe: eine wahlrechtshistorische und ideengeschichtliche Studie*, Baden-Baden, Nomos, 2000. La nota sobre la ley de derechos civiles de 1964 está tomada de manera más o menos textual de David Brady, John Ferejohn y Jeremy Pope, «Congress and civil rights policy: an examination of endogenous preferences», en Ira Katznelson y Barry Weingast (comps.), *Preferences and Situations: Points of Intersection between Historical and Rational Choice Institutionalism*, Nueva York, Russell Sage, 2005, págs. 62-87. Se hallará un examen de la votación estratégica en el Capítulo 2 de David Austen-Smith y Jeffrey S. Banks, *Positive Political Theory II: Strategy and Structure*, Ann Arbor, University of Michigan Press, 2005. Gerry Mackie, *Democracy Defended*, Cambridge y Nueva York, Cambridge University Press, 2003, contiene extensos análisis de las preferencias sociales cíclicas y la idea de que casi todos los presuntos ejemplos cíclicos en las legislaturas se basan en lecturas defectuosas de las pruebas. El ejemplo del aeropuerto de Oslo se ha tomado de Aanund Hylland, «The Condorcet paradox in theory and practice», en Jon Elster y otros, (comps.), *Understanding Choice, Explaining Behavior: Essays in Honour of Ole-Jørgen Skog*, Oslo, Oslo Academic Press, 2006, págs. 127-151. El ejemplo de la desmovilización de los soldados estadounidenses proviene de Samuel Stouffer (comp.), *The American Soldier*, Princeton, Princeton University Press, 1949. El trabajo fundacional sobre la negociación es Thomas C. Schelling, *The Strategy of Conflict*, Cambridge (Massachusetts), Harvard University Press, 1960 [*La estrategia del conflicto*, Madrid, Tecnos, 1964]. Una obra clásica sobre la negociación en acción es Howard Raiffa, *The Art and Science of Negotiation*, Cambridge (Massachusetts), Belknap Press of Harvard University Press, 1982 [*El arte y la ciencia de la negociación*, México, Fondo de Cultura

Económica, 1991]. Mi *The Cement of Society: A Study of Social Order*, Cambridge y Nueva York, Cambridge University Press, 1989 [*El cemento de la sociedad: las paradojas del orden social*, Barcelona, Gedisa, 1997] es una introducción informal a la teoría de la negociación.

Notas al pie

[56] Para decirlo de otra manera, la opción «me quedo en casa pero todos los demás tienen que votar» no se contará entre las previstas en el referéndum.

[57] Si bien la expresión «mecanismo de agregación» suele reservarse para los procedimientos de votación, yo la utilizo aquí para denotar cualquier proceso en el cual actores que pueden tener inicialmente diferentes preferencias interactúan para llegar a una decisión que todos ellos acepten con carácter vinculante.

[58] En los primeros juicios por jurado realizados en Inglaterra, es posible que la práctica de hambrear a los jurados (u obligarlos a pagar su comida) hasta que alcanzaran una decisión unánime también haya otorgado mayor poder negociador a unos que a otros.

[59] Aun las elecciones generales pueden ofrecer campo para la negociación. Si la votación es pública, los votantes y los candidatos tal vez regateen por el precio de los votos.

[60] De manera similar, se ha argumentado que si *todas* las señales animales fueran engañosas, el engaño no tendría sentido.

[61] Recuérdese, sin embargo, que esta mejora es más probable si aúnan sus datos en bruto y no simplemente las conclusiones alcanzadas sobre la base de éstos (Capítulo 23).

[62] Otras consideraciones, sobre todo la necesidad de una gobernanza eficaz, tal vez hablen en favor de la votación mayoritaria o la votación proporcional con un umbral alto. En las elecciones a asambleas constituyentes, en las cuales la gobernanza es una consideración secundaria, hay una tendencia a elegir a los delegados por medio de la votación proporcional.

[63] Los utilitaristas suelen negar la necesidad de deducir esta implicación. Por lo común argumentan que los efectos negativos del miedo y la incertidumbre que generaría el conocimiento de que uno puede ser escogido como «donante aleatorio» harían algo más que contrarrestar los beneficios de la práctica. *Pero ¿cómo lo saben?* Sospecho que razonan hacia atrás, del carácter evidentemente inaceptable del procedimiento a la existencia de costes que lo excluyan por razones utilitarias, y no hacia adelante, de la demostración de los costes al rechazo de la práctica.

[64] La Constitución sudafricana de 1996 es una excepción parcial. La exigencia de que se aprobara por mayoría calificada se incorporó a la Constitución interina de 1993, sancionada a su vez por negociación y no por votación.

[65] En la asamblea francesa, este resultado se produjo en mayo de 1791 cuando radicales, moderados y reaccionarios aunaron fuerzas para votar una ley que prohibía la elección de los miembros de la Asamblea Constituyente para la primera legislatura ordinaria. El objetivo de los radicales era debilitar la legislatura en beneficio del club de los jacobinos; el de los reaccionarios, debilitarla en beneficio del rey. La votación fue unánime, puesto que el centro moderado, «ebrio de desinterés» (capítulo 5), votó con entusiasmo para negarse a sí mismo un papel en la futura legislatura.

[66] No se pregunta, sin embargo, si la posibilidad ocasional de incurrir en *graves* errores no podría ser más seria que la parcialidad *sistemática* de un régimen no democrático. Si bien Tocqueville afirmaba que debido a su situación geográfica favorable los Estados Unidos podían permitirse cometer errores, esto podría no ser válido para otros países.

[67] Suponiendo, en contra de lo que se sostiene en el párrafo siguiente, que la probabilidad de que cada votante acierte no se ve afectada por un incremento en la cantidad de votantes.

[68] La posibilidad de que la mayoría acierte también aumenta si se exige una mayoría calificada, digamos de un 60%. En ese caso, sin embargo, podría llegarse a un «jurado en desacuerdo» en el que ni la culpa ni la inocencia del acusado obtienen la mayoría requerida.

[69] También se podría tratar de garantizar la existencia de las condiciones del teorema de Condorcet aumentando la probabilidad de que las creencias de los votantes sean de hecho independientes entre sí. Desde este punto de vista, tal vez tendría sentido la propuesta de Rousseau de proscribir la discusión con anterioridad a la deliberación. Al mismo tiempo, si la deliberación mejora la calidad de las creencias, no puede constituir una objeción el que también las haga menos independientes unas de otras. Las condiciones del teorema son suficientes para que la votación por mayoría produzca un buen resultado, pero no necesarias.

[70] Como la decisión misma de votar es irracional (Capítulo 12), cabría preguntarse si los ciudadanos no podrían también invertir irracionalmente en información acerca de las cuestiones en juego. En el presente contexto, no obstante, la cuestión relevante es si los ciudadanos invierten más cuando el derecho al sufragio es limitado que cuando es amplio, así como más votantes pueden acudir a las urnas cuando consideran reñido el resultado de la elección.

[71] En abstracto, el óptimo podría estar en uno de los extremos: sea un solo individuo o todas las personas adultas. Bajo supuestos razonables, es más probable que sea un «máximo interior». Si el tamaño óptimo es pequeño, uno podría escoger a los votantes al azar entre los ciudadanos en general, para asegurarse de que no representarán intereses sectarios. En esta perspectiva, el voto sería una *función* más que un *derecho*.

[72] Un ejemplo más reciente: cuando en 1964 se debatió en la Cámara de Representantes de los Estados Unidos el proyecto de ley de derechos civiles, el presidente de la Comisión de Procedimiento (Howard Smith) presentó lo que a su juicio era una enmienda mortal: propuso que la ley prohibiera la discriminación por motivos de sexo, por añadidura a los motivos de raza y origen nacional. Los líderes de la Cámara, que compartían la convicción de Smith de que la enmienda podía invalidar el proyecto, intentaron conseguir que los demócratas de inclinación más progresista votaran contra la propuesta, cosa que, según algunos relatos, muchos de éstos habrían estado personalmente dispuestos a hacer siempre que esa votación se realizara en privado. Pero las galerías estaban colmadas de grupos de defensa de los intereses de las mujeres que habían concurrido para observar el comportamiento en votaciones no registradas (anónimas) y, en esas circunstancias, los miembros progresistas no querían acompañar a sus líderes y rechazar la enmienda de Smith. Tal como resultaron las cosas, dicha enmienda no condujo al rechazo del proyecto sino, antes bien, a la inclusión del género como un grupo de las clases protegidas por ley.

[73] Más técnicamente: en sistemas de votación no dictatoriales y no probabilísticos, la expresión de las verdaderas preferencias no es en general una estrategia dominante. Sin embargo, es posible idear sistemas en los cuales la sinceridad sea un *equilibrio*, en el sentido de ser la mejor respuesta a la sinceridad de otros. Los detalles son complicados y no reflejan características de los sistemas de votación del mundo real.

[74] Razón por la cual el utilitarismo clásico, que nos indica maximizar la suma total de utilidades individuales, es difícil de aplicar en la práctica. Este problema no debería llevarnos a pensar que nunca tiene sentido comparar grados de bienestar entre las personas. Tanto las instituciones como los individuos llevan a cabo esas comparaciones como asunto de rutina, y lo hacen con gran seguridad. No obstante, como las comparaciones se basan en intuiciones y no en mediciones, habrá muchos casos en los que no logren producir resultados inequívocos.

[75] Como todas se preferían al statu quo, existía la certeza de que se aprobaría una de ellas.

[76] Los autores del estudio del que tomo estos datos señalaron que «era difícil esperar un alto grado de consistencia interna en elecciones hipotéticas tan intrincadas», lo cual sugiere que el problema tenía que ver con clasificaciones individualmente inconsistentes. Si las mayorías hubieran sumado más del 200%, esa sugerencia habría estado justificada. Como sólo sumaron el 167%, es muy posible que las clasificaciones fueran individualmente consistentes y, pese a ello, dieran origen a una intransitividad colectiva. El estudio se publicó en 1949, dos años antes del señero trabajo de Kenneth Arrow sobre la agregación de preferencias y las inconsistencias a las que es vulnerable.

[77] Esta formulación es un tanto engañosa. En un juego de negociación como el mostrado en la Figura 20.2, el acuerdo se alcanzará de inmediato sobre la base de la retroinducción, en la cual influye la velocidad de descuento temporal de las partes. La negociación es «virtual», no real. No obstante, sigue siendo cierto que la parte más impaciente obtendrá menos, porque le importa más conseguirlo enseguida.

Capítulo 26
Organizaciones e instituciones

Organizaciones e instituciones

Entre los ejemplos de organizaciones se cuentan el supermercado local, General Electric, el Congreso de los Estados Unidos, la Iglesia Católica, el Tribunal Constitucional alemán, Francia y la Organización de las Naciones Unidas. Entre los ejemplos de instituciones podemos mencionar la familia, las constituciones, el papel moneda, la economía de mercado y la justicia transicional. Organizaciones e instituciones pueden relacionarse como «muestra» y «tipo», esto es, un ejemplo de un concepto y el concepto mismo. El mercado bursátil es una institución (un tipo), la Bolsa de Nueva York, una organización (una muestra de ese tipo). El Estado nación es una institución, Francia es una organización. Sin embargo, algunas organizaciones, como las Naciones Unidas, no son muestra de ninguna institución. A la inversa, algunas instituciones, como la economía de mercado, no tienen muestras organizacionales. Con todo, los mercados especiales, como el de los permisos de contaminación, pueden ser organizaciones. Aunque me ocuparé sobre todo de estas últimas, se las comprende mejor contra el telón de fondo de las instituciones.

Las organizaciones son *actores colectivos*, definidos por su capacidad para la toma centralizada de decisiones. En algunos casos, su creación puede situarse en un momento específico, como el 4 de julio de 1776 para los Estados Unidos o el 24 de octubre de 1945 para las Naciones Unidas. En otros, como Francia o la Iglesia Católica, es más apropiado decir que las organizaciones *evolucionaron* o *surgieron* y no que se crearon. Si bien puede

ser arduo determinar con precisión el momento exacto en que una agregación escasamente estructurada de individuos se cristaliza en un actor colectivo con una identidad perdurable a lo largo del tiempo, habrá un punto en que podamos decir que ha sucedido.

Las instituciones son *maneras de hacer las cosas* que surgen o evolucionan poco a poco, aunque una vez que han tomado una forma definitiva las organizaciones las pueden adoptar e implementar de manera consciente. Considérense, por ejemplo, las transiciones de 1989 y 1990 en Europa oriental. Las nuevas democracias adoptaron la institución de la economía de mercado y crearon bolsas y otras organizaciones para implementarla. En sus constituciones, organizaron la institución de los tribunales constitucionales dotados de la facultad de derogar leyes anticonstitucionales. En la relación con sus predecesores autocráticos, se valieron de las instituciones de la justicia transicional.

Estas tres instituciones: la economía de mercado, los tribunales constitucionales y la justicia transicional, no siempre existieron. Aunque Adam Smith no estaba necesariamente equivocado cuando afirmaba la existencia, como un hecho universal, de «cierta propensión de la naturaleza humana […] a permutar, cambiar y negociar una cosa por otra», otros pueden poner freno a esa propensión. En particular, la puede bloquear la desconfianza que caracteriza la «guerra de todos contra todos» en el estado de naturaleza de Hobbes. Para superar esa desconfianza, acaso hagan falta instituciones. En condiciones primitivas, «permutar y cambiar» puede ser demasiado riesgoso, si el hombre fuerte sencillamente toma las posesiones del débil sin ofrecer nada a cambio. Es mejor, entonces, ocultar nuestras posesiones a las miradas ajenas. Aun cuando ese problema básico se supere mediante la vigilancia, las actitudes aprovechadas y la desconfianza pueden ser impedimentos al funcionamiento eficiente de los mercados. Si tengo una buena idea sobre la fabricación más barata de determinado producto, quizá me muestre reacio a desarrollarlo si sé que otros pueden apoderarse de él sin incurrir en los costes del desarrollo. La institución del sistema de patentes es una respuesta a ese problema. Si los prestatarios quieren usar su casa o su tierra como garantía, los prestamistas no se dejarán ver hasta cerciorarse de que la propiedad no está hipotecada varias veces. Para darles seguridades, el Estado tiene que crear un catastro que ellos puedan consultar antes de ofrecer el préstamo. Estas instituciones, que los nuevos Estados pueden ahora adoptar directamente en formas muy elaboradas, tardaron siglos en aparecer.

Acaso hoy parezca obvio que si un país tiene una constitución, también necesita un mecanismo político para invalidar las leyes anticonstitucionales. Por lo común, esta tarea («revisión judicial») corresponderá a un tribunal, ya sea uno mixto, encargado también de otras tareas, como en los Estados Unidos, o un tribunal constitucional especializado, como en el continente europeo. Históricamente, sin embargo, las constituciones existieron mucho antes de la revisión judicial.[78] Si bien la Constitución de los Estados Unidos creó una Corte Suprema que estaba autorizada a derogar las leyes estaduales, no se la autorizó en forma expresa a declarar la inconstitucionalidad de las leyes federales. Cuando en 1803 la Corte se arrogó esa facultad en el caso *Marbury v. Madison*, se trató esencialmente de un acto de fuerza. En 1854, la Corte Suprema noruega también decidió tomar a su cargo el ejercicio de la revisión judicial, que la constitución del país no menciona. En 1971, el Concejo Constitucional francés se autoatribuyó, contra las intenciones expresas de los redactores de la constitución en vigencia, la facultad de derogar las leyes que violaran la Declaración de los Derechos del Hombre y el Ciudadano de 1789. La Corte Suprema israelí también se ha atribuido el derecho a derogar la legislación violatoria de las Leyes Fundamentales del país, con lo cual convirtió su constitución parcial en una constitución con todas las de la ley. Aunque cada una de esas decisiones fue polémica en su época, todas, salvo la última mencionada, han terminado por transformarse en características indiscutidas del paisaje institucional.

En nuestros días, existe la presunción de que los dirigentes de regímenes represivos tienen que rendir cuentas ante la justicia cuando caen o dan un paso al costado. Si las nuevas autoridades nacionales no la toman por iniciativa propia, la comunidad internacional puede hacerse cargo de la tarea (Ruanda y la ex Yugoslavia), ejercer presión sobre el régimen sucesor (Indonesia) o crear tribunales híbridos con jueces tanto nacionales como internacionales (Sierra Leona y Camboya). Hoy existe algo semejante a una institución de justicia transicional. Sin embargo, si nos remontamos en el tiempo, las ejecuciones extrajudiciales, con finalidades de venganza o inhabilitación, han sido siempre el modo típico de encargarse de los líderes de regímenes represivos. En 1814 y 1815, la decisión de los aliados de exilar a Napoleón en vez de ejecutarlo fue un paso importante en el proceso de «detener la mano de la venganza». Al final de la Segunda Guerra Mundial, muchos líderes aliados querían que los principales criminales nazis fueran fusilados como forajidos donde se encontraran, en vez de someterlos a juicio. En definitiva, no obstante, los juicios de Núremberg mostraron un respeto casi ejemplar por la necesidad de

condenar sobre la base de pruebas y cumplir las exigencias del debido proceso. De hecho, algunos de los acusados fueron absueltos. En algunos países ocupados por los alemanes hubo elementos de justicia popular «salvaje» en 1944-1945, pero el imperio de la ley no tardó en reemplazarlos. Cuando los regímenes comunistas se derrumbaron cuarenta y cuatro años después, hubo una coincidencia general en el sentido de que, en el tratamiento de sus fechorías, no debían emularse sus prácticas ilegales.

El problema del monitoreo

Una organización puede tener *miembros* o *empleados*. Los miembros también pueden ser empleados, como en el caso de las cooperativas de trabajadores. Los miembros interactúan en forma horizontal, a través de los procesos de la discusión, la negociación y la votación que he examinado en el capítulo anterior. La relación vertical entre los empleados y sus superiores es de diferente carácter. Para simplificar, supongamos que la organización tiene un solo ejecutivo («el director») y muchos empleados («los agentes»). Se plantea un *problema de director y agentes* cuando, como sucede con frecuencia, los intereses de uno y otros son diferentes. A los trabajadores quizá les interese un ritmo laboral moderado, mientras que el administrador gerente tal vez desee que se esfuercen más. Si este último trata de hacer que ambos intereses coincidan pagando a los trabajadores en función del esfuerzo, es posible que le resulte difícil o costoso monitorear el empeño que ponen en trabajar. Tradicionalmente, ésa era la tarea del capataz, pero ¿cómo puede el administrador estar seguro de que éste no pide sobornos a los trabajadores o se vale de su autoridad para promover sus propios fines económicos o sexuales? El lugar de trabajo del siglo XIX se caracterizaba a menudo como «la tiranía del capataz». En casos semejantes, ¿quién custodiará a los custodios? Como veremos a la brevedad, estos problemas también pueden surgir cuando el director y los agentes coinciden, tal cual ocurre en las cooperativas de trabajadores.

De manera similar, la cabeza de un organismo estatal está interesada en que los empleados sean honestos, es decir, que no acepten ni soliciten sobornos al público. También tiene interés en la eficiencia, de modo que las dimensiones del sector público se mantengan en un mínimo posible. Los empleados pueden tener intereses opuestos en ambos aspectos. Si sólo los mueve su egoísmo económico, aceptarán sobornos cuando puedan hacerlo impunemente. A raíz de su interés en el poder, los agentes también tienen un

incentivo para inflar el tamaño de sus departamentos y multiplicar el número de subordinados. Una vez más, el monitoreo tal vez sea muy arduo. El director quizás atrape de tanto en tanto a un agente en el momento de aceptar un soborno, pero en general no puede confiar en ese método. Tal vez procure reducir las oportunidades de corrupción y apele para ello, por ejemplo, al llamado a licitaciones competitivas para la suscripción de contratos públicos, pero esta precaución no sirve si los agentes redactan los contratos a medida con el fin de favorecer a determinados proveedores. Como los agentes suelen tener un cuasimonopolio de la información, es posible que los directores sean incapaces de decir qué solicitudes de nuevos ingresos se justifican y cuáles no.

Los subordinados no son los únicos que pueden tener incentivos que discrepen con la organización. Algunos presidentes de universidades estadounidenses han tenido que dar un paso al costado luego de que se descubriera que se asignaban enormes salarios o hacían redecorar sus casas a expensas de la organización. Un vicepresidente estadounidense (Spiro Agnew) tuvo que renunciar cuando se reveló su proceder corrupto. La cleptocracia, el gobierno de los ladrones, se ha convertido en un fenómeno sumamente familiar en el mundo entero. Aunque en cierto sentido son directores, esos dirigentes también pueden ser sometidos a un monitoreo. Sin embargo, los supervisores (votantes, miembros de juntas directivas, accionistas, el Banco Mundial o el Fondo Monetario Internacional) han sido con frecuencia notoriamente ineptos para regular el comportamiento de los directores generales o los jefes de Estado. Como en otros casos, quizá carezcan de la *información* necesaria para corregir los excesos o del *incentivo* para hacerlo.

En las cooperativas de trabajadores, puede surgir un conflicto entre su papel de directores y su papel de agentes. En una alocución ante el Congreso de Ciencias Sociales en 1863, sir James Kay Shuttleworth dijo, con referencia a los molinos algodoneros cooperativos de Lancashire:

> [Entonces] se planteó la gran pregunta: ¿qué beneficios deben tener los accionistas de este molino, más allá de las ganancias ordinarias? En la práctica, la primera idea que se expuso en tales sociedades fue que debía darse preferencia a las familias de los accionistas en la selección de los trabajadores del molino. […] Quien habla había sido testigo en su propiedad del fracaso de una de estas inquietudes. Existía el deseo de introducir en la industria el principio de cooperación hasta cierto punto, a saber, que los accionistas tuvieran la prioridad para emplear a sus familiares en los molinos. El efecto inmediato de esta medida fue que, en lugar de generar una disciplina más estricta y una estrecha atención al funcionamiento de las maquinarias, que tan necesaria era en los molinos algodoneros (y quien habla podría mencionar que el rigor en la disciplina de un regimiento era inferior al de un molino), en sus reuniones trimestrales o semestrales, los trabajadores expusieran

> las quejas más enfadosas contra los capataces, y que un capataz que se había atrevido a echar a un trabajador que era accionista, corriera el peligro extremo de ser despedido en la siguiente reunión.[79]

Otro problema frecuente en las cooperativas se debe a su renuencia a suspender personal en momentos de baja demanda. Al comentar la ruina de la Wolverhampton Plate-Locksmiths en 1878, un contemporáneo escribió que

> si los negocios hubiesen estado a cargo de un productor privado, éste probablemente habría despedido a los trabajadores para quienes, a raíz de la caída de la demanda de fijaciones de placa, no pudiera encontrar una tarea rentable, y se hubiera consagrado a desarrollar las actividades restantes. Pero esto habría implicado en el cuerpo de trabajadores que formaban la sociedad un autosacrificio para el cual no estaban preparados. Decidieron, en cambio, trabajar por acciones, con la esperanza de que la demanda resurgiera. Como no sucedió así antes de que sus recursos se agotaran, sufrieron inevitablemente un quebranto. Las deudas se multiplicaron sobre sus espaldas, y los mejores trabajadores desertaron.

Soluciones al problema del director y los agentes

La holgazanería, la proliferación de nuevos empleados, la corrupción y problemas similares pueden ser contrarrestados mediante la intervención, ya sobre los incentivos de los agentes, ya sobre sus oportunidades (Capítulo 9). Esta última solución es difícil de implementar. Para ser eficaz, un agente necesita cierta independencia y libertad de acción. Los esclavos se han utilizado contadas veces en ocupaciones que exigieran aplicación y cuidado. Tal vez se demuestre la imposibilidad de estructurar la situación de manera tal que el agente tenga la libertad de perseguir las metas del director y, al mismo tiempo, sea literalmente incapaz de promover las propias. Quizá podamos tratar de acercarnos a ese objetivo si hacemos que los decisores ejerzan sus cargos por un período lo bastante breve para que sea difícil sobornarlos. El sistema de jurados y el colegio electoral de los Estados Unidos (en su concepción original) se han justificado con estos argumentos (entre otros). Se supone asimismo que el mandato breve de los funcionarios elegidos (a menudo combinado con la imposibilidad de reelegirlos) y la rotación frecuente de los funcionarios designados reducen las oportunidades de corrupción. Sin embargo, estas últimas prácticas tienen serios costes en materia de eficiencia. Cuando los funcionarios logran por fin familiarizarse lo suficiente con sus tareas para cumplirlas bien, tal vez deban dejar sus cargos.

Es posible que la intervención sobre los incentivos parezca más prometedora. Al recompensar el empeño en el trabajo o imponer multas a los holgazanes, por ejemplo, podemos armonizar los incentivos de los agentes con los del director. En ocupaciones en las cuales el esfuerzo de cada uno

puede medirse sobre la base de la producción individual, el pago a destajo tal vez resuelva el problema. A veces, no obstante, la tecnología sólo permite a la empresa medir la producción de un grupo de trabajadores. Además, el uso del pago a destajo de manera individual puede ser contraproducente. Si se paga a cada integrante de un equipo de fútbol (o a sus atacantes) en función de la cantidad de goles que hace, el equipo en su conjunto tendrá un desempeño menos eficiente. Un buen jugador de equipo es aquel que pasa el balón a otro que está mejor ubicado para anotar. La empresa podría entonces tratar de resolver el problema estableciendo un sistema de bonificaciones por equipo, con el fin de hacer que cada trabajador tuviera un incentivo para monitorear a los demás.

No obstante, si, como suele suceder, los costes del monitoreo son superiores a los beneficios que corresponden a la persona responsable de ejercerlo, esta última carecerá de incentivos *materiales* para llevarlo adelante. Un jugador de fútbol quizá compruebe que no merece la pena tomarse la molestia (acaso considerable) de reprender a uno de sus compañeros por su comportamiento de «estrella». Si bien es posible que surja una norma social contra la holgazanería, al respecto las pruebas empíricas son ambiguas. Es más probable que normas de ese tipo se desarrollen en pequeños grupos, debido a que sus integrantes pueden observarse con más facilidad unos a otros y a que cada uno de ellos sufre un perjuicio más severo si cualquier otro se comporta como un aprovechado. Con todo, aun en este caso las normas tal vez no surjan de manera completamente espontánea (Capítulo 22). La norma contra el comportamiento de «estrella» o el egoísmo en el fútbol quizá deba mucho a las estrategias del entrenador, que puede castigar a los jugadores que se sobrestiman y subestiman al equipo. Hay, sin embargo, pruebas anecdóticas de que, para castigar a quienes no pasan el balón a los que están mejor ubicados, los jugadores suelen, a su vez, no pasárselo.

La cuestión de quién custodiará al custodio, o supervisará a los supervisores, podría en teoría resolverse mediante algo semejante a un *sistema circular*, en el que cada persona integrante de una cadena tuviera un incentivo material directo para castigar a quien está a su izquierda, dado que si no lo hace, será castigada por quien está a su derecha. En la práctica, no parece haber ningún dispositivo de ese tipo. Los jugadores de fútbol pasan el balón a quienes lo pasan a quienes no lo hacen. Con todo, la tentación de aprovecharse podría ser mitigada por la *confianza*. Si las empresas ofrecen salarios por encima de la tasa vigente, es posible que los trabajadores (o algunos de ellos) respondan con un esfuerzo adicional. El obstáculo es que la

empresa tal vez no tenga incentivos para establecer ese plan. Aunque la producción total (y por lo tanto los ingresos totales) se incremente, otro tanto hará el conjunto de la masa salarial. Si el resultado neto consiste en menores ganancias, a la empresa le convendrá más ofrecer un salario competitivo.

La organización de la desconfianza

Las constituciones políticas aspiran con frecuencia a ser un sistema circular de pesos y contrapesos, en el cual el desempeño de cada institución es monitoreado por otra. Esta aspiración no se realiza en todas las constituciones, y acaso no se realice en ninguna. La maquinaria constitucional suele apoyarse en un «móvil inmóvil» o un «controlador sin control», una parte que monitorea el desempeño de otras sin estar sometida a su vez a ningún monitoreo. En los Estados Unidos, la Corte Suprema ocupa esa posición. En otros países, la tarea está a cargo del parlamento. En Rumanía, por ejemplo, el parlamento puede anular fallos del tribunal constitucional si éste deroga sus leyes (un legado de la era comunista, cuando aquél era formalmente omnipotente, pero impotente en los hechos). También en Francia el parlamento fue el móvil inmóvil hasta 1971. No obstante, al menos dentro de la tradición anglonorteamericana, el diseño constitucional se apoya en un principio formulado por David Hume: «Es […] una máxima política justa suponer que todos los hombres son tunantes». Desde ese punto de vista, las constituciones son una forma de desconfianza organizada.

La Constitución de los Estados Unidos es tal vez la expresión más elaborada de esta actitud. Entre los rasgos llamativos de los debates de la Convención Federal de 1787 y el documento producido por ésta se cuentan la preocupación constante por la posibilidad de corrupción, sobornos y amenazas y la necesidad de cerrar todas las grietas por las cuales esos procederes pudieran colarse. Para mencionar un ejemplo, la Comisión Especial sobre la representación que informó el 9 de julio recomendó que, «como la situación presente de los estados probablemente se modifique tanto en lo tocante a la riqueza como a la cantidad de sus habitantes, se autorice a la legislatura, de tiempo en tiempo, a aumentar su número de representantes». En respuesta a esa recomendación, Edmund Randolph se manifestó «temeroso de que, como el número no se ha de modificar hasta que la Legislatura Nacional así lo quiera, nunca falte el pretexto para posponer las modificaciones y mantener el poder en manos de quienes lo poseen». Con un razonamiento similar, George Mason argumentó que, «a causa de la naturaleza del hombre, podemos estar

seguros de que quienes tienen el poder en sus manos no lo cederán mientras puedan conservarlo». En términos más débiles, deberíamos insistir con Hume en el *supuesto* de que quienes están en el poder tratarán de retenerlo. Aunque algunos políticos tal vez trabajen por el bien común (y todos sostengan hacerlo), no debemos malgastar la confianza. Para decirlo de manera más contundente, es preciso suponer que quienes tienen poder lo utilizarán para conservarlo y, de ser posible, *ampliarlo*. Desde esta óptica, las constituciones son dispositivos para impedirles hacerlo. Y aunque no es lo único que hacen, a menudo es una de las cosas que aspiran a hacer.

La capacidad de este marco normativo para *explicar* la promulgación de constituciones reales depende de la intención de los redactores de buscar el bien común y no promover sus intereses o pasiones particulares. Como señalé hacia el final del Capítulo 13, no puede darse por descontado, a buen seguro, que estén exentos de estas últimas motivaciones. Sin embargo, la presión normativa para actuar *como si* los motivara la razón –la fuerza civilizadora de la hipocresía– puede ser más fuerte en una asamblea constituyente que en una legislatura ordinaria. Además, como he sostenido en el capítulo anterior, la influencia del egoísmo en las cuestiones que deben decidirse suele ser menor en el ámbito constituyente. Por esa razón, las asambleas constituyentes que sesionan a puertas cerradas tal vez logren reducir el efecto distorsivo de la pasión sin pagar el coste de una negociación excesivamente basada en los intereses.[80]

Con esta salvedad, querría presentar un ejemplo esquemático de cómo podría la desconfianza organizada incorporarse a una serie de disposiciones interconectadas. Muchas de las características se encuentran en sistemas reales. Otras se incluyen sobre todo con el fin de ilustrar hasta dónde podría llegarse para cerrar todas las grietas posibles.

1. Para impedir que el Gobierno *ejerza una justicia política*, la constitución debe requerir jurados pa los juicios penales o disponer que los jueces sean asignados al azar a los casos (o seleccionados p algún otro procedimiento mecánico). En ausencia de cláusulas de este tipo, el gobierno podría enjuiciar a sus opositores políticos y asegurar su condena mediante la elección a dedo del juez.
2. Para impedir que el Gobierno *manipule el sistema electoral*, la ley electoral debe incorporarse has en sus detalles más específicos a la constitución y no quedar librada a la determinación estatutaria La reformulación de los distritos electorales para tomar en cuenta los cambios demográficos podr encararse por medio de una fórmula mecánica o bien confiarse a una comisión independiente. La historia norteamericana y francesa reciente muestra muchos casos en que las autoridades en ejercicio han modificado las reglas electorales para permanecer en el poder.
3. Para impedir que el Gobierno *manipule el flujo de información*, la radio y la televisión públicas deben ser manejadas por comisiones independientes («modelo británico»), no designadas por las autoridades nacionales («modelo francés»). Los medios privados también deben ser protegidos. E preciso encargar a comisiones independientes la asignación de las frecuencias. Es menester prohil

al gobierno el racionamiento de recursos que son vitales para los medios como el papel y la tinta d imprenta, o confiar esas decisiones a una «junta de racionamiento independiente». Se ha demostrado que la combinación de elecciones y medios no manipulados es sumamente eficaz para impedir hambrunas y otros desastres. Si quienes están tentados de actuar en contra del interés público saben que sus actividades quedarán expuestas a la vista del electorado, que podrá entonce desalojarlos del poder, es menos probable que muestren ese comportamiento.

4. Para impedir al Gobierno la *manipulación de la política monetaria* como instrumento para la reelección, dicha política debe confiarse a un banco central independiente. Una solución alternativ pero menos atractiva, consistiría en dejar que las decisiones en esa materia quedaran en manos de un programa informático diseñado por tres economistas especialistas en el tema elegidos al azar e los cinco mejores departamentos universitarios de economía del país; el programa, por otra parte, debería revisarse cada cinco años.
5. Para impedir que autoridades dogmáticas o ideológicas de los bancos centrales *implementen políticas monetarias desastrosas*, debería existir la posibilidad de destituirlas por decisión de una mayoría calificada (dos tercios o tres cuartos) del parlamento. Su cargo debería tener una duración limitada (a diferencia de los ex gobernadores del Banco Central de Italia), y sería necesario que só pudiese ejercerse una sola vez (a diferencia del gobernador italiano actual).
6. Para impedir al Gobierno *la manipulación de la información estadística* como un instrumento par la reelección, la recolección de datos debería quedar en manos de un organismo estadístico independiente. En Gran Bretaña, con el fin de promover la confianza en las cifras oficiales, el ministro de Hacienda ha dispuesto que la Oficina de Estadísticas Nacionales sea independiente de Gobierno (*Financial Times*, 2 de diciembre de 2005). Un titular del *The New York Times* del 22 de septiembre de 2002 decía: «Algunos expertos temen influencia política sobre los organismos de recolección de datos del delito». También se han denunciado presiones políticas sobre las estadísticas laborales («La Oficina de Estadísticas Laborales debería ser investigada», *New York Post*, 29 de septiembre de 2000). En casos más sutiles, también puede estar en juego la elección d un modelo de pronósticos econométricos. Es evidente que cuanto menores sean el delito y el desempleo en opinión de la gente, y mejores las perspectivas de crecimiento, mayores serán las posibilidades de reelección para los titulares de los cargos electivos. Además de prohibir al Gobierno la utilización de la política monetaria para manipular la realidad, habría que impedirle usar las estadísticas para manipular las creencias sobre la realidad.
7. Para impedir al Gobierno *hambrear a la oposición*, los principales partidos políticos deberían recibir subsidios en función de la proporción de votos y bancas en el parlamento. La constitución debería establecer el monto de esos subsidios, que podrían atarse al precio del oro.[81]
8. Para impedir al Gobierno *la promulgación de leyes en beneficio propio*, la constitución podría estipular que ciertas categorías de leyes sólo deben entrar en vigor algún tiempo después (un año o más) de ser sancionadas. Por medio de este mecanismo se crearía un velo artificial de ignorancia, detrás del cual aun las partes sólo preocupadas por la defensa de sus intereses podrían verse en la necesidad de tomar medidas imparciales.
9. Para impedir al Gobierno *eludir esas restricciones* mediante el uso de su mayoría parlamentaria co el propósito de reformar la constitución, las reformas deberían exigir la existencia de mayorías calificadas, plazos o ambas cosas. La cláusula misma de reforma debería ser inmune a las reforma
10. Para impedir al Gobierno *ignorar esas restricciones*, como sucedía anteriormente en Francia, la constitución debería establecer la revisión judicial a cargo de un Tribunal Constitucional o una Corte Suprema independientes.
11. Para impedir al Gobierno *manipular la revisión judicial* por medio del aumento de la cantidad de jueces del Tribunal Constitucional o la Corte Suprema, ese número no debería establecerse a través de leyes, como ocurre en los Estados Unidos y varios otros países, sino quedar estipulado en la constitución.
12. Para impedir que jueces de inclinaciones dogmáticas o ideológicas del Tribunal Constitucional o l

Corte Suprema *ignoren las grandes mayorías populares*, debe existir la posibilidad de destituirlos por decisión de una mayoría calificada (dos tercios o tres cuartos) de los legisladores. A diferencia de lo que sucede con los magistrados de la Corte Suprema de los Estados Unidos, debería limitarse la duración de sus mandatos. Tampoco tendrían que ser reelegibles.

Como lo indica este ejemplo, una constitución puede ser una maquinaria intrincada. De faltar una de las piezas interconectadas, otras pueden llegar a ser ineficaces o nocivas (véase la discusión ulterior). Al mismo tiempo, existe el peligro real de que la constitución sea tan rígida que impida al sistema político responder a las emergencias. Llegado el caso, las restricciones constitucionales en materia de política monetaria o leyes electorales podrían ser desastrosas. En la Convención Federal celebrada en Filadelfia, George Mason señaló que «aunque profesaba un odio mortal por el papel moneda, como no podía prever todas las emergencias, no estaba dispuesto a atar las manos de la legislatura. Quien habla ha advertido que, de haber existido esa prohibición, la última guerra no podría haberse librado». De manera similar, cuando en 1946 el parlamento italiano decidió no incluir la estabilidad monetaria en la constitución, una de las objeciones se refería a la necesidad de que el Gobierno tuviera libertad para actuar en tiempos de guerra. En 1945, De Gaulle manipuló el sistema electoral francés para mantener a los comunistas al margen del poder. Muchos años después, en un comentario sobre una propuesta para incorporar la ley electoral a la constitución, dijo que «uno nunca sabe lo que puede pasar. Algún día podría volver a haber razones para apelar a la votación proporcional en bien del interés nacional, como en 1945. No debemos atarnos las manos». En retrospectiva, Francia debería agradecer que la constitución no incluyera la ley electoral.

El problema puede sintetizarse con dos célebres observaciones sobre la Constitución estadounidense. En una referencia a la ley de 1871 consagrada al Ku Klux Klan, John Potter Stockton dijo que «las constituciones son cadenas con las que los hombres se atan en sus momentos de cordura para no morir por mano suicida en el día de su frenesí». En 1949, Robert Jackson afirmó que «la Declaración de Derechos no es un pacto suicida». No hay una manera sencilla de garantizar que una constitución prevista como un dispositivo de prevención del suicidio no se transforme algún día en un pacto suicida.

El problema de la segunda opción

Dejemos de lado esta cuestión y supongamos ahora que si las doce condiciones que he enumerado (u otras del mismo tipo general) se cumplen, el

sistema político será en algún sentido (que para nuestra finalidad presente no necesita ser muy preciso) *óptimo*. En el mundo real, es improbable que se alcance este resultado. Es difícil que todo esté bien. Sin embargo, podríamos aspirar a *acercarnos* lo más posible al óptimo, sobre la base del supuesto aparentemente razonable de que, cuanto más se cumplan las condiciones de la optimalidad, más cerca estaremos de ella. Este supuesto es falso. En circunstancias muy generales, no es cierto que una situación en la cual se cumplen muchas (no todas) de las condiciones para un óptimo sea necesariamente –y ni siquiera probablemente– superior a otra en que se cumplen menos condiciones. Con referencia a las doce disposiciones antes mencionadas, no está claro a priori que la revisión judicial sin un control sobre los jueces sea mejor que ninguna revisión judicial, o que los bancos centrales independientes sin control sobre los banqueros sean mejores que un sistema que deja la política monetaria en manos del gobierno.

El examen que Tocqueville hace del Antiguo Régimen en Francia puede leerse en esta perspectiva. Ese sistema se caracterizó por una serie de rasgos que estarían ausentes en una sociedad bien ordenada. La administración real tenía facultades amplias, mal definidas y arbitrarias. La venalidad de los cargos hacía imposible la existencia de una burocracia racional. La obstrucción de los *parlements*, tribunales muy politizados que actuaban sobre todo en beneficio propio, dificultaba la implementación de políticas coherentes. No obstante, sostenía Tocqueville, dado el primero de estos rasgos, la presencia de los otros dos era, en realidad, beneficiosa:

> El gobierno, en su deseo de convertirlo todo en dinero, había empezado por poner en venta la mayor parte de los cargos públicos, y de ese modo se había privado de la facultad de otorgarlos y revocarlos a voluntad. Así, una de sus pasiones había interferido en sumo grado en el éxito de la otra: su codicia había contrarrestado su ambición. En consecuencia, para actuar se veía constantemente reducido a utilizar instrumentos que no había confeccionado y que no podía destruir. Por ello, con frecuencia la ejecución de sus deseos más absolutos era débil. Esta extravagante y defectuosa constitución de las funciones públicas ocupaba el lugar de cualquier tipo de garantía política contra la omnipotencia del gobierno central. Era una extraña y desvencijada suerte de dique que dividía los poderes del gobierno y mitigaba su influencia. […] La intervención irregular de los tribunales en el gobierno, que a menudo desorganizaba la administración adecuada de los asuntos, servía así por momentos para salvaguardar la libertad: era un gran mal que limitaba otro aún más grande.

En su análisis del derecho inglés en el siglo XVIII, James Fitzjames Stephen señaló que el énfasis excesivo en detalles técnicos «morigeraba efectivamente, aunque de manera irracional y caprichosa, la excesiva severidad de las viejas leyes penales». En un comentario sobre el derecho penal islámico, sostuvo, en forma similar, que «mitigaba la extravagante

dureza de sus disposiciones mediante reglas de prueba que, en la práctica, excluían la posibilidad de llevarlas a cabo». De manera análoga, el Imperio otomano, la Rusia zarista, la Italia de Mussolini y la España de Franco han sido caracterizados como un «despotismo moderado por la incompetencia». En efecto, comparados con la eficiencia implacable de la Alemania nazi, esos regímenes eran benignos.

Tocqueville también observaba que, cuando los funcionarios del Antiguo Régimen procuraron reemplazar la odiada *corvée* (trabajos forzados en las carreteras) por un impuesto que se utilizaría para mantener los caminos, terminaron abandonando el proyecto por temor a que, «una vez establecido ese fondo, nada impidiera al tesoro desviarlo para sus propios fines, con el resultado de que, pronto, los contribuyentes se vieran gravados con un nuevo impuesto y sometidos, asimismo, a la obligación de los trabajos forzados». Este fenómeno es muy común. Cuando los políticos proponen que los hospitales públicos den prioridad en los turnos a los individuos a quienes podría ayudarse a volver a integrar la fuerza de trabajo, suelen prometer canalizar las ganancias así generadas hacia esos mismos hospitales, con el objetivo de beneficiar también a otros pacientes. Por lo común, los administradores hospitalarios son escépticos con respecto a esas propuestas, pues sospechan que las ganancias se destinarán simplemente a llenar las arcas oficiales. En California, las propuestas de dar agua a los establecimientos agrícolas capaces de asignarle el destino más adecuado, con el fin de generar un excedente que podría utilizarse para mejorar la provisión a otros agricultores, han sido recibidas con similar escepticismo. Una situación con dos elementos por debajo del óptimo –un sistema ineficiente de prioridades para asignar bienes escasos y un gobierno que no respeta sus promesas (o las de sus predecesores)– tal vez sea mejor que otra en la cual se suprime el primero de esos elementos.

Estos casos se asemejan en parte a lo que ocurre cuando los decisores políticos eliminan animales que son una molestia para las poblaciones humanas, para comprobar a continuación que los organismos que aquéllos mantenían bajo control provocan un perjuicio aun más grande. Así, tras decidir eliminar a los gorriones porque comían grano, Mao Tse-tung tuvo que importarlos de la Unión Soviética debido a la expansión de las plagas que esos pájaros mantenían a raya, con catastróficas consecuencias ecológicas. No menos que los ecosistemas, las sociedades pueden tener rasgos aparentemente absurdos o nocivos cuya eliminación tal vez produzca males aún mayores. Es

probable que esta razón, entre otras, haya impulsado a Edmund Burke y sus seguidores a criticar con tanto vigor la planificación institucional racionalista.

Esta línea de razonamiento puede llevarse demasiado lejos. Cuando los trabajadores conquistaron el derecho a votar, los adversarios de la medida argumentaron que la supresión de una injusticia cometida con ellos desestabilizaría la sociedad en todo tipo de aspectos y crearía una injusticia aún más grande con los ricos, cuyos caudales correrían el riesgo de ser confiscados por los pobres. Este argumento pone en un pie de igualdad dos clases de consideraciones que no tienen el mismo peso. Pero como este libro se dedica a la explicación de los fenómenos sociales y no a su evaluación desde un punto de vista normativo, no profundizaré en la cuestión.

* * *

NOTA BIBLIOGRÁFICA

Tomo la distinción entre organizaciones e instituciones de Claus Offe, «Political institutions and social power: conceptual explorations», en Ian Shapiro, Stephen Skowronek y Daniel Galvin (comps.), *Rethinking Political Institutions: The Art of the State*, Nueva York, New York University Press, 2006, págs. 9-31. El «ascenso del mercado» es un tema central en Douglass C. North, *Structure and Change in Economic History*, Nueva York, Norton, 1981 [*Estructura y cambio en la historia económica*, Madrid, Alianza, 1984]. La aparición de la revisión judicial es el tema de Eivind Smith (comp.), *Constitutional Justice Under Old Constitutions*, La Haya y Boston, Kluwer Law International, 1995. Sobre el surgimiento de la justicia transicional, véase Gary J. Bass, *Stay the Hand of Vengeance: The Politics of War Crimes Tribunals*, Cambridge (Massachusetts), Harvard University Press, 2001. Una útil antología dedicada a lo que yo llamo de manera un tanto idiosincrásica organización es Claude Ménard (comp.), *The Political Economy of Institutions*, Cheltenham, Edward Elgar, 2004. El problema de la armonización de los incentivos individuales y organizacionales es el tema estudiado por Jean-Jacques Laffont y Jean Tirole, *A Theory of Incentives in Procurement and Regulation*, Cambridge (Massachusetts), The MIT Press, 1994. Arnold J. Heidenheimer, Michael Johnston y Victor T. LeVine (comps.), *Political Corruption: A Handbook*, New Brunswick (Nueva Jersey), Transaction Publishers, 1989, es un manual general sobre la corrupción. Las referencias a las cooperativas inglesas del siglo XIX se extraen de Benjamin Jones, *Co-Operative Production*, Oxford, Clarendon Press, 1894, y Nueva

York, Kelley, 1968. La idea de un «sistema circular» de monitoreo mutuo se examina en Eugene Kandel y Edward Lazear, «Peer pressure and partnership», *Journal of Political Economy*, 100(4), 1992, págs. 801-817, y en Randall Calvert, «Rational actors, equilibrium, and social institutions», en Jack Knight e Itai Sened (comps.), *Explaining Social Institutions*, Ann Arbor, University of Michigan Press, 1995, págs. 57-93, reeditado en C. Ménard (comp.), *The Political Economy of Institutions*, *op. cit*. La importancia relativa de la confianza y los incentivos en las empresas se discute en Ernst Fehr y Armin Falk, «Psychological foundation of incentives», *European Economic Review*, 46(4-5), 2002, págs. 687-724. La influencia de Hume en los orígenes de la Constitución estadounidense se explora en Morton White, *Philosophy*, The Federalist*, and the Constitution*, Nueva York, Oxford University Press, 1987. La cita de De Gaulle proviene de Alain Peyrefitte, *C'était De Gaulle*, vol. 1, *La France redevient la France*, París, Fayard, 1994, pág. 452. La idea de la segunda opción fue expuesta por Richard G. Lipsey y Kelvin Lancaster, «The general theory of the second best», *Review of Economic Studies*, 24, 1956, págs. 11-32. Examino los análisis de Tocqueville consagrados al Antiguo Régimen en «Tocqueville on 1789: preconditions, precipitants, and triggers», en Cheryl Welch (comp.), *The Cambridge Companion to Tocqueville*, Cambridge y Nueva York, Cambridge University Press, 2006, págs. 49-80. Los pasajes citados de James Fitzjames Stephen pertenecen a su *History of the Criminal Law of England*, Londres, Macmillan, 1883, y Buffalo (Nueva York), Hein, 1964, vol. 1, pág. 284, y vol. 3, pág. 293. La referencia a Mao Tse-tung es de Jung Chang y Jon Halliday, *Mao: The Unknown Story*, Nueva York, Alfred Knopf, 2005, págs. 430-431 [*Mao: la historia desconocida*, Madrid, Taurus, 2005]. Expongo mis puntos de vista sobre el fundamento normativo del cambio institucional en el Capítulo 4 de *Solomonic Judgments: Studies in the Limitations of Rationality*, Cambridge y Nueva York, Cambridge University Press, 1989 [*Juicios salomónicos: las limitaciones de la racionalidad como principio de decisión*, Barcelona, Gedisa, 1991].

Notas al pie

[78] Holanda aún no tiene revisión judicial. La revisión judicial sueca está limitada por la cláusula constitucional de que una ley tiene que ser «evidentemente» inconstitucional para que el tribunal la derogue. Es de presumir que dichas leyes tendrían, por lo pronto, muy pocas posibilidades de sancionarse.

[79] En 1875, el director general de la Ouseborn Co-Operative Engine Works Company se quejó de situaciones similares para explicar el fracaso de la empresa: «Las reglas horarias de Ouseborn, incluida la de que cada día tuviera sus propias horas extras, las excesivas restricciones al trabajo de una sección de nuestros hombres, la pequeña proporción de nuestros aprendices, las frecuentes discusiones y disputas, la cantidad desmesurada de horas extras requeridas para permitirnos cumplir con nuestros contratos y, en general, la falta de disciplina resultante de la impresión de que cada hombre era su propio jefe: todo contribuyó a incrementar los costes de producción».

[80] No me siento del todo cómodo con esta concepción negativa de la pasión. Hegel sostenía que «nada grande se ha realizado en el mundo sin pasión». Tocqueville escribió que, en tiempos de crisis, la gente se sitúa o bien por debajo o bien por encima de su nivel normal. Por otra parte, la imparcialidad no excluye la pasión, como lo demostraron en 1789 los redactores constitucionales franceses que estaban «ebrios de desinterés». Uno de los más imponentes, el conde de Clermont-Tonnerre, dijo que «la anarquía es una etapa alarmante pero necesaria, y el único momento en que es posible establecer un nuevo orden de cosas. Las medidas uniformes no pueden tomarse en épocas de calma». Sin embargo, aun cuando el entusiasmo pueda resultar en fines admirables, tal vez debilite la claridad de pensamiento indispensable para realizarlos.

[81] Un deseo correspondiente de hacer que la constitución fuera lo más explícita y menos manipulable posible subyacía a la intención de Madison de fijar los salarios de los jueces federales «utilizando como patrón el trigo o alguna otra cosa de valor permanente».

Conclusión

¿Es posible la ciencia social?

¿A qué se considera ciencia?

La respuesta a la pregunta de si una ciencia social es posible, o acaso ya existe, depende de los criterios utilizados para definir la ciencia. Un criterio externo podría ser el siguiente: una disciplina se ha convertido en ciencia cuando 1) entre quienes la practican hay en cualquier momento una coincidencia general en torno de lo que es verdadero, lo que es falso, lo que es conjetural y lo que es desconocido dentro de su ámbito; 2) hay un proceso de avance acumulativo en virtud del cual las teorías y explicaciones, cuando se desechan, se desechan para siempre; 3) los principales conceptos y teorías pueden expresarse en términos lo bastante claros y explícitos para que los entienda cualquier persona dispuesta a invertir tiempo y esfuerzo en ello, y 4) los «clásicos» de la disciplina son leídos principalmente por historiadores de las ciencias. Como dijo Alfred Whitehead: «Una ciencia que vacila en olvidar a sus fundadores está perdida».

Esta descripción aspira a abarcar de manera aproximada el estatus de las ciencias naturales contemporáneas. Ninguna de estas disciplinas cumple literal y plenamente el criterio 1. Existen controversias, por ejemplo, acerca de la teoría de las cuerdas, el equilibrio puntuado o la importancia relativa de lo innato y lo adquirido en el desarrollo humano. No obstante, es fácil que la profundidad de la discrepancia sea exagerada por quienes no pertenecen a la comunidad científica pertinente, incapaces, tal vez, de apreciar la enorme cantidad de conocimientos compartidos que constituyen el telón de fondo de la controversia, y que en ocasiones confunden las exploraciones tentativas con tesis definitivas. El criterio 2 suele cumplirse. En la historia de las ciencias

naturales hay pocas inversiones, si las hay: ningún neonewtoniano que se rebele contra Einstein y ningún neolamarckiano que se rebele contra Darwin. La transformación de lo que quizás hayan sido teorías algo oscuras o torpes al formularse por primera vez en material de libro de texto de fácil asimilación da testimonio del cumplimiento general del criterio 3. Con escasas excepciones, sobre todo en biología (Darwin, D'Arcy Wentworth Thompson, Claude Bernard), el criterio 4 también se cumple.

Apelaré ahora a estos criterios para proponer una evaluación de las ciencias sociales blandas, cualitativas y cuantitativas. Luego de algunas breves observaciones desdeñosas sobre la ciencia social blanda, expondré un argumento en favor de la ciencia social cualitativa. Y, de manera más polémica, presentaré otro argumento igualmente favorable a la ciencia social cuantitativa, al menos en algunas de sus formas más destacadas. El proponente de la ciencia social cualitativa, por consiguiente, se encuentra en la incómoda situación de librar una guerra en dos frentes, con el riesgo constante de que uno de sus adversarios lo acuse de aliarse con el otro.

La ciencia social blanda

Las ciencias sociales «blandas» tienen más en común con ciertas formas de la crítica literaria (o con la literatura) que con las investigaciones cualitativas empíricas. Muchos autores, el más eficaz de los cuales quizás haya sido Alan Sokal, han expuesto el posmodernismo, la teoría poscolonial, las teorías de los subalternos, deconstruccionismo, el psicoanálisis kleiniano o lacaniano y teorías similares como los oscurantismos que efectivamente son. Según he señalado en una entrevista, la falta de un lenguaje compartido para la discusión racional hace imposible criticar frontalmente a estos seudoteóricos. En cambio, es menester hacerlos tirar piedras contra su propio tejado, como Sokal hizo con suma eficacia al lograr que una de sus revistas publicara un artículo presentado por él y dedicado a la hermenéutica de la gravedad cuántica, colmado de una jerga sin sentido pero de sonido rimbombante.

Aunque los integrantes de estas camarillas desarrollen en su medio un estilo de discusión que sugiere cierto grado de cumplimiento del criterio 1, dicho estilo no equivale sino a una forma de seudointersubjetividad en la cual los reflejos verbales son una mascarada del acuerdo racional. En cuanto a los criterios 2 a 4, no se cumplen nunca o casi nunca. Al parecer, ninguna idea se desecha para siempre. Lo más llamativo es, tal vez, el incumplimiento del

criterio 3. Derrida quizá deslumbre con su lenguaje, pero sus «enseñanzas» no son *enseñables*. No hay manuales «prácticos» de deconstruccionismo (aunque haya muchas investigaciones o «mapeos» y «negociaciones»), sobre todo porque quienes lo practican tienden a proceder mediante la insinuación y las preguntas retóricas en vez de arriesgar el pescuezo y hacer aseveraciones definidas. Lo más cercano a un manual tal vez sea *Postmodern Pooh*, el conjunto de parodias de Frederick Crew. Con respecto al criterio 4, el culto de los antepasados, con referencias y exégesis obligatorias, parece imponerse. (En este párrafo, es obvio que he predicado a los conversos.)

La ciencia social cualitativa

Dentro de la ciencia social cualitativa incluyo el grueso de los escritos históricos, así como las obras que toman la forma de «estudios de casos» y no de estudios de «un gran número». Creo que, para cualquier especialista en ciencias sociales, la mejor formación consiste en leer historia en abundancia y con profundidad, escogiendo las obras por la cualidad intrínseca del argumento y no por la importancia o relevancia del tema. Menciono aquí algunos modelos: James Fitzjames Stephen, *History of the Criminal Law of England*; Edward P. Thompson, *The Making of the English Working Class* [*La formación de la clase obrera en Inglaterra*]; Geoffrey Ernest Maurice de Ste. Croix, *The Class Struggle in the Ancient Greek World* [*La lucha de clases en el mundo griego antiguo*]; Joseph Levenson, *Confucian China and Its Modern Fate*; Paul Veyne, *Le Pain et le cirque*; Georges Lefebvre, *La Grande peur de 1789* [*El gran pánico de 1789: la Revolución francesa y los campesinos*]; Keith Thomas, *Religion and the Decline of Magic*; Tocqueville, *L'Ancien régime et la Révolution* [*El Antiguo Régimen y la Revolución*]; Max Weber, *Agrarverhältnisse im Altertum*; Gordon Wood, *The Radicalism of the American Revolution*; Jean Egret, *La Pré-révolution française*; Denis Crouzet, *Les Guerriers de Dieu*, o Martin Ostwald, *From Popular Sovereignty to the Sovereignty of Law* (he arriesgado un poco el pescuezo al incluir algunos clásicos todavía no reconocidos). Lo que estos autores y otros de su talla tienen en común es el hecho de combinar una total autoridad en cuestiones fácticas con la atención a las generalizaciones potenciales y los potenciales contraejemplos que las refutan. En virtud de su conocimiento, pueden detectar tanto el «detalle revelador» como la «anomalía sólida» y proporcionar estímulo y un cotejo con la realidad a los aspirantes a generalistas.

Lo mismo vale para los autores de «estudios de casos», entre los cuales *La democracia en América*, de Tocqueville, sigue siendo uno de los más grandes. Aunque no entra del todo en esta categoría, también incluiría a Joseph Schumpeter y su *Capitalism, Socialism, and Democracy* [*Capitalismo, socialismo y democracia*]. Un candidato aparentemente excéntrico, pero a mi juicio ineludible, es *Travels in France* de Arthur Young, que abarca los años 1787, 1788 y 1789. Éstos son «retratos de carácter» de sociedades o regímenes enteros, y todos ellos adoptan una perspectiva comparativa. *La Société féodal* [*La sociedad feudal*] de Marc Bloch también pertenece a esta categoría. *Cumbres abismales*, de Aleksandr Zinoviev, no es exactamente un retrato de carácter del comunismo soviético, sino una caricatura en el buen sentido de la palabra, pues elimina los aspectos que no son esenciales y aísla rasgos centrales mediante su exageración. Una versión plenamente documentada es la de Stéphane Courtois y otros, *Le Livre noir du communisme* [*El libro negro del comunismo*]. La trilogía de Richard J. Evans sobre el Tercer Reich* promete hacer con respecto al régimen específico del nazismo lo que Robert Paxton hizo en *The Anatomy of Fascism* [*Anatomía del fascismo*] con referencia a este régimen más genérico. *Mussolini* y *Mussolini's Trial*, de Richard Bosworth, si se leen en conjunción con los libros de Evans, proporcionan visiones sorprendentes de la diferencia entre un régimen cuya malignidad, si bien real, era en gran medida de inferior calidad, y otro que era maligno hasta la médula.

El análisis histórico y los estudios de casos no se limitan a suministrar materia prima para la generalización: a menudo contienen en sí mismos generalizaciones implícitas cuya revelación el autor deja en manos del lector. Tocqueville, por ejemplo, era un maestro de la generalización latente. Debido a una combinación de la arrogancia del historiador y del aristócrata, no se dignó a sacar a la luz su andamiaje teórico, y pretendió no hacer otra cosa que avanzar paso a paso. Si bien *La democracia en América* se apoya constantemente en el esquema de los deseos y las oportunidades para explicar el comportamiento, nunca se lo enuncia con todas las letras. Y aunque *El Antiguo Régimen y la Revolución* apela en abundancia a la distinción entre la envidia y el odio en el análisis de las relaciones de las clases, los lectores deben conjeturar por sí mismos las causas y efectos sistemáticamente diferentes de esas dos emociones.

Bajo el encabezado de la ciencia social cualitativa también incluyo escritos que *explicitan* esas generalizaciones. Para examinarlas, no recurriré a una enumeración de autores u obras, sino que volveré a los enigmas que

presenté en la introducción. Es muy importante destacar que las «respuestas» están sujetas a la segunda salvedad mencionada en la Introducción: sólo satisfacen la condición mínima de implicar el *explanandum*. En algunos casos, arriesgaría el pescuezo y diría que la explicación que propongo es probablemente la correcta, pero en otros no me animaría a hacerlo.

PRIMERA PARTE: LA MENTE

- ¿Por qué algunos jugadores creen que cuando el rojo ha salido cinco veces seguidas, tiene más probabilidades que el negro de volver a salir la vez siguiente? *Respuesta*: porque están sujetos a la heurística de la disponibilidad.
- ¿Por qué otros jugadores creen que cuando el rojo ha salido cinco veces seguidas, el negro tiene más probabilidades de salir la vez siguiente? *Respuesta*: porque están sujetos a la heurística de la representatividad.
- ¿Por qué las preferencias cambian a veces con el mero paso del tiempo? *Respuesta*: o bien debido al descuento hiperbólico o debido a la breve vida media de las emociones que las generan.
- ¿Por qué tanta gente que parece creer en la vida después de la muerte quiere que ésta llegue lo más tarde posible? *Respuesta*: porque su creencia no es del tipo que sirve como premisa para la acción, y se limita a proporcionarles cierta paz espiritual cada vez que piensan en la muerte.
- ¿Por qué la gente es reacia a reconocerse y reconocer ante los demás, que es envidiosa? *Respuesta*: porque se preocupa por la imagen de sí misma y porque la envidia, en la mayoría de las sociedades, está cerca del fondo en la jerarquía normativa de las motivaciones.
- ¿Por qué la gente es reacia a reconocerse y reconocer ante los demás, que es ignorante? *Respuesta*: porque los seres humanos son animales buscadores de patrones y porque la confesión de ignorancia en cuestiones importantes provoca malestar psíquico.
- ¿Por qué, entre los conversos del siglo XVI al calvinismo, la creencia de que la gente estaba predestinada o bien al cielo o bien al infierno inducía más paz espiritual que la creencia de que era posible alcanzar la salvación mediante las buenas obras? *Respuesta*: porque el pensamiento mágico los llevaba a creer que, al unirse a la Iglesia, podían hacer que Dios los incluyera entre los elegidos.
- ¿Por qué es verdad (a veces) que «quien ha ofendido, no puede perdonar»? *Respuesta*: porque los individuos orgullosos son tan renuentes a admitir que han actuado mal que inventan razones por las cuales la persona ofendida merecía y sigue mereciendo su destino.
- ¿Por qué en algunas culturas la vergüenza es más importante que la culpa? *Respuesta*: porque una sociedad que no ha conceptualizado la culpa también exhibirá menos comportamientos culposos.
- ¿Por qué la victoria francesa en la Copa del Mundo de fútbol de 1998 generó tanta alegría en el país, y por qué el hecho de que el equipo francés no pasara de la primera fase en 2002 causó tanta desazón? *Respuesta*: porque la sorpresa magnifica las emociones tanto positivas como negativas.
- ¿Por qué las mujeres a menudo sienten vergüenza después de haber sido violadas? *Respuesta*: porque comparten con frecuencia la actitud de «culpar a la víctima» que se deduce de la creencia de que el mundo es fundamentalmente un lugar justo.
- ¿Por qué los rituales humillantes de iniciación generan más lealtad y no menos al grupo en el cual uno es iniciado? *Respuesta*: porque la creencia de que uno ha sufrido un gran dolor por nada generaría una disonancia cognitiva.

SEGUNDA PARTE: ACCIÓN

- ¿Por qué hoy son más numerosos que hace veinte años los espectáculos de Broadway que suscitan ovaciones de pie del público? *Respuesta*: porque el auditorio necesita sentir que recibe valor a cambio de su dinero.
- ¿Por qué los castigos incrementan, en vez de disminuir, la frecuencia del comportamiento al que apuntan? *Respuesta*: porque el odio generado por el uso de castigos crueles puede contrarrestar el miedo.
- ¿Por qué la gente es reacia a romper reglas autoimpuestas, aun en los casos en que tiene poco sentido respetarlas? *Respuesta*: porque teme que una sola excepción cause la desintegración de la regla.
- ¿Por qué el modelo de venganza de alguna gente es «dos ojos por uno» y no «ojo por ojo»? *Respuesta*: debido a la aversión a la pérdida.
- ¿Por qué el rendimiento a largo plazo de las acciones es muy superior al de los bonos (es decir, por qué el valor de las acciones no sube hasta igualar los rendimientos)? *Respuesta*: debido a una combinación de aversión a la pérdida y «miopía decisional».
- ¿Por qué los índices de suicidio disminuyen cuando los medicamentos peligrosos se venden en blísteres y no en frascos? *Respuesta*: porque muchos deseos son tan efímeros que, en el momento en que uno consigue abrir el blíster, el impulso suicida ya se ha desvanecido.
- ¿Por qué ninguno de treinta y ocho espectadores llamó a la policía cuando Kitty Genovese fue asesinada a golpes? *Respuesta*: porque cada uno de ellos creía que, como no había intervenido ningún otro, la situación no era probablemente muy grave.
- ¿Por qué algunos individuos escondieron o rescataron a judíos bajo el régimen nazi? *Respuesta*: porque se les pidió que lo hicieran y la vergüenza les impedía negarse.
- ¿Por qué el presidente Chirac convocó a elecciones anticipadas en 1997, con el único resultado de perder su mayoría parlamentaria? *Respuesta*: porque no previó que los votantes extraerían del anuncio de la convocatoria a elecciones anticipadas una información que habría de llevarlos a votar contra él.
- ¿Por qué algunos padres que se divorcian están dispuestos a compartir la custodia de los hijos cuando la solución que prefieren es la custodia exclusiva, que probablemente obtendrían si litigaran? *Respuesta*: porque sienten aversión al riesgo.
- ¿Por qué los pobres son menos propensos a emigrar? *Respuesta*: porque no pueden solventar los costes del viaje y tampoco pueden ponerse como garantes de sí mismos para obtener un préstamo.
- ¿Por qué algunas personas depositan ahorros en cuentas navideñas que no pagan intereses y no permiten retiros antes de Navidad? *Respuesta*: porque saben que si depositaran sus ahorros en una cuenta común, el descuento hiperbólico tal vez los induciría a retirarlos por anticipado.
- ¿Por qué la gente encara proyectos, como la construcción del avión Concorde, que tienen un valor esperado negativo? *Respuesta*: debido a la altivez o a la aversión a la pérdida.
- ¿Por qué, en la «justicia transicional» (situación en que los agentes de un régimen autocrático son sometidos a juicio luego de la transición a la democracia), quienes son juzgados inmediatamente después de la transición reciben sentencias más severas que quienes son juzgados más adelante? *Respuesta*: debido a la breve vida media de las emociones justicieras.
- ¿Por qué, en la obra de Shakespeare, Hamlet posterga la venganza hasta el último acto? *Respuesta*: porque está sometido a la debilidad de la voluntad y porque la tensión no puede resolverse antes del final de la pieza.

TERCERA PARTE: LECCIONES DE LAS CIENCIAS NATURALES

- ¿Por qué es mucho más probable que los padres maten a hijos adoptados o hijastros que a sus hijos biológicos? *Respuesta*: porque sólo estos últimos llevan sus genes.
- ¿Por qué es tan escaso el incesto entre hermanos, si se tienen en cuenta las tentaciones y oportunidades? *Respuesta*: porque la selección natural ha favorecido un mecanismo que inhibe el

deseo sexual por miembros de la misma edad y del sexo opuesto dentro de la misma casa.
- ¿Por qué la gente invierte su dinero en proyectos emprendidos por otros agentes, aun cuando éstos tengan la libertad de apropiarse de todos los beneficios? *Respuesta*: porque la selección grupal ha favorecido una tendencia a cooperar.
- ¿Por qué, al vengarse, la gente acepta pagar algún coste material, sin recibir beneficios de la misma especie? *Respuesta*: porque la selección grupal ha favorecido una tendencia a castigar a quienes no cooperan.
- ¿Por qué la gente se apresura a sacar conclusiones que las pruebas existentes no justifican? *Respuesta*: porque la selección natural ha favorecido la búsqueda de patrones.

CUARTA PARTE: INTERACCIÓN

- ¿Por qué los adherentes de un partido socialista a veces votan a los comunistas y de ese modo impiden la victoria de su partido? *Respuesta*: porque cada uno de ellos es víctima del «síndrome del hermano menor», que le impide advertir que otros podrían hacer lo mismo.
- ¿Por qué algunos países que se han independizado recientemente adoptan como idioma oficial el de su anterior opresor imperialista? *Respuesta*: porque en un país con muchas lenguas locales, el idioma de la potencia colonial es la segunda opción de todos.
- ¿Por qué los puestos de helados de la playa suelen estar unos al lado de otros, cuando los clientes estarían mejor atendidos, y a los vendedores no les iría peor, si estuvieran más separados entre sí? *Respuesta*: porque para cada vendedor es individualmente racional desplazarse hacia el medio, con prescindencia de lo que haga el otro.
- ¿Por qué un individuo vota en elecciones cuando es virtualmente cierto que su voto no tendrá influencia en el resultado? *Respuesta*: porque el votante es presa del pensamiento mágico o actúa en función del imperativo categórico.
- ¿Por qué los individuos económicamente exitosos de las sociedades occidentales modernas suelen ser más delgados que la persona común y corriente? *Respuesta*: porque creen (con acierto) que otras personas abrigan la creencia (falsa) de que la gente que carece de autocontrol en un ámbito probablemente tampoco lo tendrá en los demás.
- ¿Por qué la gente se abstiene de hacer transacciones que podrían mejorar la situación de todos, por ejemplo, cuando evita preguntar a la primera persona de la fila en una parada de autobús si está dispuesta a vender su lugar? *Respuesta*: porque es más reacia a las muestras abiertas que a los despliegues ocultos de desigualdad económica.
- ¿Por qué el presidente Nixon trató de mostrarse ante los soviéticos como si fuera propenso a un comportamiento irracional? *Respuesta*: porque ese comportamiento daba credibilidad a unas amenazas de destrucción mutua en las que, de lo contrario, los soviéticos tal vez no hubieran creído.
- ¿Por qué los jefes militares a veces queman los puentes (o sus propias naves)? *Respuesta*: porque esperan que su enemigo, al saber que no podrán retirarse, se abstenga de emprender un costoso combate.
- ¿Por qué la gente suele atribuir gran importancia a cuestiones formales intrínsecamente insignificantes? *Respuesta*: porque cree que alguien que se aparta de las normas no se preocupa por lo que piensen de él.
- ¿Por qué los pasajeros dan propina al conductor de un taxi y los comensales hacen lo mismo con un camarero, aun cuando se encuentren en una ciudad extranjera a la que no tienen intención de volver? *Respuesta*: porque la idea de que otros puedan pensar mal de ellos es dolorosa.
- ¿Por qué las empresas hacen extensos inventarios aunque no prevean ninguna interrupción de la producción? *Respuesta*: porque esperan que esa inversión disuada a los trabajadores de hacer huelga, y con ello se aseguran de que no haya interrupciones.

- ¿Por qué, en un grupo de estudiantes, cada uno piensa que los otros han entendido mejor que él un texto oscuro? *Respuesta*: porque cada uno de ellos es víctima de un «síndrome del hermano mayor» que lo lleva a pensar que, si los demás no buscan ayuda, no se debe a la timidez.
- ¿Por qué en muchas asambleas políticas se vota pasando lista? *Respuesta*: porque los partidarios de las medidas populares utilizan este procedimiento para disuadir a quienes, de lo contrario, podrían votar contra ellas.
- ¿Por qué el intercambio de favores es más frecuente en las legislaturas ordinarias que en las asambleas constituyentes? *Respuesta*: porque lo sostiene un comportamiento fundado en el «hoy por ti, mañana por mí», que exige una interacción constante durante períodos más prolongados.

¿Cómo se desempeña la ciencia social cualitativa en los criterios 1 a 4 que expuse al comienzo? Para tomarlos en orden inverso, diremos que los clásicos no son obsoletos. Me costaría mucho tomar en serio a quien sostuviera que hoy no merece la pena dar importancia a las obras clásicas, porque sus descubrimientos, cuando son exactos, están plenamente incorporados al pensamiento actual. Los clásicos tienen mucho más que un interés de anticuario. No afirmo, sin embargo, que un diálogo con los maestros del pasado sea la única o la mejor manera de generar nuevas ideas. Thomas Schelling, por ejemplo, no parece, al menos de manera evidente, pararse sobre los hombros de nadie. Kenneth Arrow tal vez haya redescubierto y generalizado la intuición de Condorcet, pero no hay influencia de éste sobre él. Que yo sepa, no fue el conocimiento de ningún precursor lo que dio origen a la obra de Daniel Kahneman y Amos Tversky. Cuando tuve en una ocasión la oportunidad de señalar a este último que una de sus distinciones (entre el «efecto de dotación» y el «efecto de contraste») había sido prevista por Montaigne y Hume, Tversky se limitó a replicar que lo alegraba encontrarse en tan buena compañía. Como los estudiosos que acabo de mencionar son responsables de los progresos posiblemente más decisivos en las ciencias sociales de los últimos cincuenta años, es obvio que no puede sostenerse que el diálogo con el pasado es el único camino a las nuevas ideas. Tomemos otro ejemplo: cuando Tocqueville se propuso escribir *El Antiguo Régimen y la Revolución*, se abstuvo deliberadamente, según propia confesión, de leer los escritos previos sobre el tema, y se limitó a los archivos y a la lectura de los textos contemporáneos de los acontecimientos que estudiaba. Prefería, dijo, redescubrir algunas de las verdades encontradas por otros que sufrir la influencia indebida de ellos. Si bien hay pruebas de que, en realidad, echó un vistazo a lo que se había escrito antes, su obra maestra debe en general poco a los otros.

Dicho esto, el diálogo con el pasado puede ser inmensamente fructífero, aunque sólo sea para identificar las posiciones que es menester refutar. Es difícil imaginar que no marxistas como Weber o Schumpeter pudieran haber

escrito lo que escribieron de no haber leído con detenimiento a Marx. La influencia directa o positiva también es corriente, por supuesto. Parece probable que algunas teorías recientes sobre la evolución de los sistemas de propiedad hayan sido directamente influenciadas por David Hume, en vez de ver a éste como un mero precursor. La obra de Paul Veyne sobre la psicología de la tiranía en la Antigüedad debe mucho al análisis de Hegel sobre la relación entre el amo y el esclavo. George Ainslie, que ha hecho mucho para presentar una de las intuiciones fundamentales de Freud de manera analíticamente convincente, tal vez no hubiera llegado a sus ideas de no ser por la versión incipiente y anterior de aquél. Sospecho que las *Tácticas de las asambleas legislativas* de Bentham no han sido aún objeto de una explotación suficiente. En estos casos, como en otros que tengo en mente, las ideas inspiradas por los clásicos deben sostenerse por sí mismas una vez que se llega a ellas. El buen uso de los clásicos no incluye los argumentos de autoridad.

En la mayoría de los ejemplos que he citado, la ciencia social cualitativa aprueba con todos los honores el criterio 3. A mi juicio, lo que podríamos llamar «giro analítico» de las ciencias sociales no se apoya en el uso de una metodología cuantitativa, sino en una preocupación casi obsesiva por la claridad y la precisión. (En algunos filósofos analíticos, esa preocupación puede adoptar una forma cabalmente obsesiva.) La importancia de distinguir entre los vínculos conceptuales y los vínculos causales entre los objetos de estudio recibe un reconocimiento creciente. «Contexto» se lee cada vez más como «niebla» y no como «un rico entorno». En comparación con el saber anterior, hay mucho menos esencialismo; los estudiosos se preguntan con menos frecuencia qué son «realmente» la democracia o el socialismo. Se admite que si bien el uso limita hasta cierto punto las definiciones, éstas no pretenden captar esencias subyacentes. Se trata de estipulaciones arbitrarias que sólo debemos juzgar en función de su capacidad para permitirnos dar con buenas explicaciones de fenómenos interesantes.

También hay, creo (o espero), una tendencia a apartarse de lo que podríamos llamar pensamiento «abstracto no deductivo». Para ser honesto, el razonamiento abstracto tiene que someterse a la disciplina o bien de la lógica deductiva o bien de una referencia constante a los hechos que muestran que las proposiciones abstractas tienen una aplicación y *marcan una diferencia*. En el pasado, las abstracciones a menudo cobraban vida propia. El significado de los términos podía cambiar en el transcurso de una argumentación, permitiendo inferencias inválidas. Marx, por ejemplo, «dedujo» la propiedad

privada a partir del trabajo alienado de la siguiente forma: como bajo el capitalismo el producto no «pertenece al» trabajador, en el sentido de que el trabajo carece de significado, tiene que «pertenecer a» –ser propiedad de– otra persona, el capitalista. Un efecto beneficioso de la profesionalización sin pausa de las ciencias sociales en la mayoría de las instituciones educativas de Occidente consiste en hacer que resulte mucho más difícil continuar con este tipo de cosas. Las apelaciones a la analogía también han perdido preponderancia, aunque todavía las hay. Si bien la idea de «capital humano» es una valiosa prolongación de la idea de capital físico, no puede decirse lo mismo de «capital de consumo» (Gary Becker), «capital cultural» (Pierre Bourdieu) y «capital social» (Robert Putnam). En el mejor de los casos, estas expresiones son metáforas inútiles e inofensivas; en el peor, abren infructuosos caminos de investigación y sugieren hipótesis causales falsas.

Es más arduo discernir hasta qué punto la ciencia social cualitativa cumple el criterio 2. Entre los historiadores (incluidos quienes trazan retratos de carácter de un régimen), las opiniones van y vienen. Considérese, por ejemplo, la idea de que el Terror estuvo presente en la Revolución francesa desde sus comienzos, en comparación con la tesis de que fue el resultado del ascenso evitable de Robespierre y de la huida igualmente evitable del rey, que deslegitimó la monarquía y provocó las guerras revolucionarias. Conforme a esta última concepción, si Mirabeau hubiera vivido, tal vez nada de esto habría pasado. Siempre aparece algún historiador francés que sostiene uno u otro punto de vista, pero la concepción mayoritaria es cíclica. La que hoy se desecha puede ser mañana dominante. Cuesta imaginar alguna cuestión histórica fundamental que no haya generado y no siga generando fluctuaciones similares. Es de presumir que siempre será así. Aun en los casos en que los datos están dados y es posible someterlos a una completa investigación, como sucede hasta cierto punto con la historia antigua, nuevas generalizaciones producidas por las ciencias sociales no históricas pueden llegar a confirmar algunas interpretaciones en detrimento de otras.

En las ciencias no históricas hay un progreso claro, pero no del conocimiento en sentido literal. Consideremos una vez más nuestra manera de entender por qué los agentes racionales pueden aspirar a dejar de lado algunas de sus opciones, quemando por ejemplo los puentes. La idea de que ese comportamiento aparentemente autorrestrictivo puede ser una conducta *del todo* racional contra un adversario, y no una simple protección racional contra nuestras inclinaciones irracionales, es un logro irreversible en materia de comprensión. La cantidad de episodios reales de puentes o naves quemadas

que la idea nos permite explicar es harina de otro costal. Un jefe militar podría quemar sus naves, como hizo Cortés, sencillamente para impedir que sus tropas emprendan la huida. El discernimiento de la posibilidad de mayorías cíclicas es irreversible, pero el número de casos reales a los que se aplica es posiblemente pequeño. El «descubrimiento» (una manera de denominar una intuición preanalítica común) de que las emociones tienen una vida media breve no quedará invalidado, pero en cualquier caso concreto puede llegar a haber mecanismos que, al actuar en otras direcciones, contrarresten la desintegración espontánea. Aun cuando la explicación dominante de un suceso o episodio determinado sea descartada y luego rehabilitada, los elementos básicos o los mecanismos en acción en el descarte y la rehabilitación persisten. El repertorio, o el tamaño de la caja de herramientas, no se contrae.

En lo concerniente al criterio 1, las ciencias sociales cualitativas no funcionan bien. Como ya se ha señalado, puede haber (de hecho, casi siempre hay) discrepancias dentro de una generación determinada de historiadores. Aunque no tomemos en cuenta las concepciones políticas que podrían llevar a marxistas y liberales a ver de diferente manera el Terror, hay mucho lugar para la controversia académica. La atribución de motivaciones y creencias a los actores históricos, por ejemplo, es un asunto muy delicado. El comportamiento que un historiador toma como una expresión ingenua de estados mentales será considerado estratégico por otro. Un aristócrata francés exilado durante el Terror tal vez se diga convencido que la monarquía no tardará en restaurarse, pese a lo cual alquila un piso por largo tiempo. Es probable que sus contemporáneos y el historiador lo crean menos sincero que a otro exilado que toma alquileres sucesivos por períodos breves; sin embargo, es posible que aun este último actúe de ese modo para suscitar una impresión de optimismo inconmovible. Sea como fuere, es indudable que sus enemigos afirmarán que lo hace por esa razón. A menos que el historiador tenga la fortuna de encontrar un documento de dudosos propósitos estratégicos, como un recordatorio que el actor histórico haya escrito para sí mismo, la cuestión sólo podrá resolverse a través del *juicio*. Y aunque los buenos historiadores lo son en parte porque tienen buen juicio, también pueden equivocarse.

Dentro de las ciencias sociales cualitativas no históricas también puede haber discrepancias sustanciales. Los economistas más influyentes sostienen a veces que muchos de los descubrimientos de la economía de la conducta que parecen constituir pruebas de irracionalidad se ven bajo esa luz debido a que

se obtienen en ámbitos experimentales. Por diversas razones, lo que las personas hacen en el marco artificial de un laboratorio puede no corresponder a su comportamiento espontáneo en la vida cotidiana. Con el fin de dar a los experimentos una organización que enfrente esa objeción, los economistas conductistas han elevado las apuestas, aislado a los sujetos de la presión social de otros sujetos y del experimentador y dado cabida al aprendizaje a lo largo del tiempo. También se ha planteado la objeción de que los sujetos transfieren mecánicamente respuestas que son adaptativas fuera del laboratorio al ámbito experimental, donde son desadaptativas, por ejemplo cuando utilizan estrategias del tipo de «hoy por ti, mañana por mí» en experimentos sin réplica en los que no tienen sentido. La respuesta de los economistas conductistas es que no hay nada mecánico en el comportamiento del sujeto en el laboratorio, que puede demostrarse capaz de ajustes finos basados en los incentivos. Estos debates no han perdido vigencia, y para quien no es especialista resulta difícil emitir una opinión. Si tiendo a alinearme con los economistas conductistas, tal vez lo haga en cierta medida por la razón nada intelectual de que el oponente de mi oponente es mi amigo.

Ciencia social cuantitativa

Me ocupo ahora de las ciencias sociales cuantitativas. En éstas hay tres campos: medición, análisis de datos y modelización. Los tres campos están vinculados, dado que el análisis de datos a menudo requiere mediciones (para establecer esos datos) y también necesita con frecuencia una modelización (que nos diga qué datos buscar). Me concentraré en esta última, dado que mis conocimientos sobre la medición y el análisis de datos son demasiado escasos, pero diré ante todo algunas palabras sobre éstos.

La medición de entidades como el consumo per cápita, el desempleo, la magnitud de la corrupción o la opinión del público acerca de la pena de muerte es una actividad intrínsecamente difícil. Para que las mediciones del consumo permitan comparaciones a lo largo del tiempo y entre países, es preciso considerar que, en distintos momentos y lugares, los consumidores tal vez compren diferentes artículos con el mismo ingreso, que sus ingresos difieren y que los bienes en venta pueden variar. La medición del desempleo es difícil si hay una gran economía subterránea, una abundante población estudiantil, una vasta población carcelaria o muchas personas que han abandonado la búsqueda de trabajo. Las mediciones de la corrupción se basan habitualmente en sus niveles percibidos, conforme los determinan

evaluaciones de expertos y encuestas de opinión. Por razones obvias, es arduo obtener pruebas independientes para evaluar la confiabilidad de esas fuentes. Aun en los casos en que convergen, quizá lo hagan porque están sometidas a un mismo sesgo. La medición de la opinión pública presupone la existencia de una opinión estable que puede medirse. Sin embargo, es sabido que cambios irrelevantes en las preguntas pueden implicar grandes diferencias en la respuesta. «¿Está usted a favor de A?» suscita respuestas muy diferentes de «¿Está usted a favor de A si la alternativa es B?», aun cuando sea claro que B es la única alternativa a A. Como ejemplo, sea A el ingreso de un país a la Unión Europea y B el no ingreso. Estos problemas pueden afectar de manera muy sustancial la solidez del análisis de datos.

El análisis de datos (expresión con la que aludo al análisis estadístico) es en cierto sentido el núcleo de las ciencias sociales modernas. Si nuestro interés consiste en el estudio de fenómenos sociales de gran escala, sabemos que siempre deberemos manejarnos con poblaciones que son heterogéneas en una serie de parámetros, como la salud, el ingreso, la situación familiar, las preferencias y el lugar de residencia. Cualquier predicción determinista en el sentido de que, al ser expuestos al impacto externo X (por ejemplo, un incremento de la tasa impositiva marginal), *todos* los individuos responden con el comportamiento B (por ejemplo, la reducción de su oferta laboral), está condenada al fracaso. Lo que se procura es, en cambio, determinar *la probabilidad* de que un individuo perteneciente a una categoría dada responda de cierta manera, y con ese fin se establece la correlación entre la pertenencia a dicha categoría y la respuesta. Con frecuencia, se trata de ejercicios sumamente técnicos.

También pueden ser ejercicios peligrosos, en cuanto tienen una serie de trampas en las que tal vez caigan aun los investigadores experimentados. En el Capítulo 1 y en el Capítulo 2 mencioné brevemente cinco de ellas: la explotación de datos, el ajuste de curvas, la arbitrariedad en la medición de variables dependientes o independientes, el problema de distinguir la correlación de la causación y la dificultad para identificar la dirección de la causación. La práctica común de utilizar valores «rezagados» de las variables, correlacionar el valor de una variable en el tiempo t con el valor de otra en el tiempo $t + n$, brinda más oportunidades de manipulación. Además, la heterogeneidad de las unidades de análisis puede ser un problema. En un análisis estadístico de las causas de la guerra, no es evidente que la Primera Guerra Mundial y la Segunda Guerra Mundial deban considerarse a la par con una «guerra del fútbol» entre dos países latinoamericanos. Sin embargo, tal

vez cabría cuestionar también la práctica de eliminar los casos anómalos como «valores atípicos», a fin de mejorar el ajuste. Un problema adicional es el del sesgo de la selección, que puede surgir, por ejemplo, si las encuestas de opinión se basan en entrevistas telefónicas, lo cual excluye a los individuos que no tienen teléfono o que, por alguna razón, no contestan los llamados.

Para un no especialista como yo es imposible decir con cuánta frecuencia los científicos sociales caen en estas y otras trampas del análisis de datos. El patrón oro tiene que ser la predicción o «posdicción», que predice una parte de las observaciones a partir del análisis realizado sobre otra parte. El mero descubrimiento de un patrón en los datos es menos convincente, dadas las incesantes posibilidades de manipulación y maniobra. También éstas podrían en principio reducirse si los investigadores suscribieran un compromiso público sobre la utilización de hipótesis y procedimientos con anterioridad al análisis concreto. Que yo sepa, esto sucede muy pocas veces.

La modelización es un ejercicio deductivo que comienza con supuestos y termina con predicciones. En las ciencias sociales, la modelización basada en la elección racional es la forma más común, pero dista de ser la única. Es cada vez más frecuente que los científicos sociales propongan modelos evolutivos que no presuponen respuestas racionales y ni siquiera intencionales a los sucesos. Estos científicos también han propuesto modelos de comportamiento más mecánicos, esto es, modelos sin fundamentos ni justificaciones claras. El físico inglés Lewis Richardson, por ejemplo, ha elaborado un modelo de la carrera armamentística basado en dos ecuaciones diferenciales lineales que no tiene un basamento psicológico claro. Por mi parte, me limito a la modelización de los fenómenos sociales basada en la elección racional.

Dicha modelización está muy difundida en la ciencia económica, en la cual se aplica hoy mucho más allá de los campos tradicionales del comportamiento de consumidores y productores. Existen, entre otras, la economía del suicidio, la economía de la concurrencia a las iglesias y, en términos más generales, el análisis económico de cualquier actividad que de una u otra manera involucre una *elección*. Hay una tendencia a eliminar la cuña conceptual que separa la acción intencional de la acción racional: si existe una oportunidad de elección, se supone que se la ejercerá racionalmente. En las ciencias políticas, la teoría de la elección racional es indudablemente dominante, en cuanto implica mayor prestigio y salarios más altos y es cada vez más habitual su carácter obligatorio en los programas de posgrado de las instituciones de elite (de los Estados Unidos). Las aplicaciones se concentran con más frecuencia en los políticos (y a veces en

los partidos e incluso los Estados) que en los votantes o los ciudadanos. En sociología, la minoría que aplica la teoría de la elección racional a los problemas tradicionales de la disciplina ha sido desplazada hasta cierto punto por economistas con mayor dominio de las herramientas (el estudio de la confianza es una excepción importante). En antropología social, rama que (en los Estados Unidos) se ha encaminado hacia las ciencias sociales blandas, esta teoría no ha conquistado apoyo.

¿Es esto ciencia? Me gustaría examinar ante todo el primer criterio. Podemos dar por sentado que la modelización es correcta, en el sentido de que las conclusiones se deducen de las premisas. ¿Son ciertas las conclusiones? Antes de abordar esta cuestión, quiero señalar que no está claro que siempre *pretendan* serlo, es decir, que correspondan al mundo real. Antes bien, a veces representan una forma de ciencia ficción: un análisis de la acción y la interacción de agentes idealmente racionales, que nunca han existido ni existirán. No puede decirse que el análisis de formas cada vez más refinadas de equilibrios estratégicos, por ejemplo, esté motivado por un deseo de explicar o pronosticar el comportamiento de individuos reales. Por el contrario, la motivación parece ser de orden estético. Dos de los más consumados teóricos de los equilibrios, Reinhart Selten y Ariel Rubinstein, han señalado con toda claridad que no creen que sus modelos tengan nada que decir sobre el mundo real. Cuando abordan el funcionamiento de éste, utilizan alguna variedad de economía de la conducta o racionalidad limitada. Para mencionar otro ejemplo, la teoría de la elección social (el estudio axiomático de los mecanismos de votación) llegó a ser en un momento tan complicada desde el punto de vista matemático y tan evidentemente irrelevante para el estudio de la política real, que una de las revistas más destacadas de economía, *Econometrica*, suspendió la publicación de artículos relacionados con ese campo.

Una cuestión interesante en la psicología y la sociología de la ciencia es cuántos son los *practicantes secretos* de la ciencia ficción económica, que se ocultan a sí mismos u ocultan a otros el hecho de que eso es, en efecto, lo que practican. La invención de modelos matemáticos ingeniosos es una actividad bien paga, pero con excepción de personas como Selten y Rubinstein, el pago sólo se efectivizará si también puede afirmarse que la actividad es relevante; de allí el incentivo para el engaño o el autoengaño. Acaso parezca que el planteamiento de esta cuestión está fuera de los límites del discurso académico, pero no veo por qué tenga que estarlo. Más allá de cierto punto, las normas académicas de cortesía deberían dejarse a un lado, algo que

también se dijo en *Sobre la libertad* de John Stuart Mill, que desde su aparición ha sido la biblia en materia de libertad intelectual. Ya he abandonado toda simulación de cortesía para con los oscurantistas de la ciencia social blanda y no veo motivo alguno para tratar de diferente manera el *oscurantismo duro*.

Carezco de pruebas directas del engaño o el autoengaño, de modo que tengo que proceder de otra manera. Trataré de mostrar, entonces, que lo que hacen los profesionales de la elección racional está a menudo tan alejado de la realidad que cuesta tomar en serio sus pretensiones de estar comprometidos con el mundo. A riesgo de repetir lo que he dicho en capítulos anteriores, querría intentar sintetizar mi argumento en diez puntos.

1. Muchos modelos atribuyen a los agentes motivaciones que éstos manifiestamente no tienen. El supuesto habitual del descuento temporal exponencial es un ejemplo.
2. Algunos modelos atribuyen a los agentes motivaciones para las que no se aduce prueba alguna, ta como una forma particular de la función de utilidad. Se puede, desde luego, estipular una función fin de demostrar teoremas de la posibilidad («un consumidor racional con utilidad marginal decreciente *podría* hacer X»), pero no probar nada acerca del mundo. En muchas situaciones, también debe demostrarse que el supuesto de que los agentes son egoístas es cierto, y no limitarse suponerlo.
3. La mayor parte de los modelos ignoran los avances en la comprensión de la elección asociada cor la teoría de la perspectiva. Como la teoría se enuncia en términos de ganancias o pérdidas a partir una línea de base dada, no analiza el comportamiento desde el punto de vista de la maximización una función objetiva definida más allá de los resultados. El formato de la maximización quizá sea indispensable para la manipulación matemática, pero irrelevante para gran parte del comportamiento real.
4. Algunos modelos adjudican a los agentes capacidades cognitivas que éstos tal vez posean pero no ejercen. Es posible, por ejemplo, que no realicen espontáneamente el razonamiento de la retroinducción que es necesario en muchas aplicaciones importantes.
5. Muchos modelos adjudican a los agentes capacidades cognitivas que éstos manifiestamente no tienen. La observación es tan trivial que resulta casi embarazosa: ¿cómo puede un economista suponer que un agente tiene la aptitud de llevar a cabo cálculos que él (el economista) necesita exponer en muchas páginas de apéndices extremadamente técnicos? Sentimos la fuerte tentación decir: «¡Vamos! ¡Ponga los pies sobre la tierra!»
6. Muchos modelos atribuyen intenciones sobre la base del interés objetivo del agente en una situaci específica, sin detenerse a preguntar si no hay tal vez otras motivaciones –emociones o normas sociales– en juego. Si tratamos de explicar el comportamiento en situaciones turbulentas o conflictivas como las revoluciones, por ejemplo, la adjudicación de un motivo de egoísmo racional de largo plazo (reducido al valor presente por el descuento exponencial) parece casi ridículamente inadecuada. Otros modelos atribuyen intenciones sobre la base de los resultados reales, y de ese modo pasan por alto la posibilidad tanto de cálculos erróneos como de jugadas calculadas.
7. Muchos modelos omiten tomar en cuenta la niebla de incertidumbre que rodea la mayoría de las decisiones importantes, sobre todo cuando las consecuencias pertinentes se sitúan en un futuro distante. Los intentos de superar este problema mediante la asignación de probabilidades subjetiva a los resultados suelen ser arbitrarios, como sucede cuando los investigadores apelan al principio razón insuficiente para estipular una distribución uniforme de las probabilidades.
8. Algunos modelos estipulan colectividades (clases o Estados) como si fueran agentes unitarios, sin

abordar (en el caso de las clases) el problema de los aprovechados ni (en el caso de los Estados) la estructura de la toma colectiva de decisiones.

9. Algunos modelos suponen que las desviaciones con respecto a la racionalidad 1) son temporarias, bien 2) se compensarán en el agregado. En cuanto al inciso 1, las desviaciones temporarias pueden ejercer efectos duraderos si la situación tiene una estructura de «trampa de langostas». Por lo que refiere al inciso 2, la economía de la conducta ha mostrado que, habida cuenta de que muchas desviaciones son sistemáticas y no aleatorias, no hay razón para esperar que desaparezcan en el agregado. El «enigma de la prima de equilibrio» es un ejemplo.
10. Muchos modelos se conforman con explicar el comportamiento mediante el argumento de que se trata de un equilibrio en un juego, sin dar un paso más y mostrar por qué, en situaciones con muchos equilibrios, se ha realizado ése en particular. Por otra parte, contadas veces hacen la pausa necesaria para preguntarse si el mundo no cambia con tanta rapidez que los equilibrios no tienen tiempo suficiente para establecerse.

He indicado algunas respuestas típicas a estas objeciones. Una consiste en mencionar el problema hermenéutico de establecer motivaciones de una manera no circular. Coincido en que esta dificultad es seria, pero no insuperable; y aun si lo fuera, no justificaría la atribución gratuita de motivaciones. Otra consiste en proponer el reemplazo de la racionalidad concreta (de agentes reales y vivos) por una «racionalidad del como si». He sostenido que ésta es una estratagema que sólo puede funcionar si se cumple una de dos condiciones. Por un lado, quizá podríamos señalar un mecanismo capaz de remedar la racionalidad *de principio a fin*, hasta los apéndices matemáticos. El problema es que nadie ha dado con un mecanismo de esas características. El recurso a la selección es poco más que un gesto, dada la enorme discrepancia entre la tosquedad y la imperfección de los mecanismos de selección social y los procesos ultrasofisticados de razonamiento atribuidos a los agentes. Por otro lado, tal vez fuéramos capaces de señalar predicciones tan irresistiblemente precisas que nos veríamos obligados a aceptar la teoría aunque no entendiéramos cómo funciona. El problema es que las ciencias sociales no producen ninguna predicción que se acerque siquiera remotamente a esa condición.

De hecho, si examinamos con más detenimiento el cumplimiento del criterio 1, observamos la coexistencia en el tiempo de diferentes teorías con explicaciones y predicciones muy distintas. En este aspecto, sobresale en particular la macroeconomía. He perdido el rastro del estatus actual del keynesianismo, el monetarismo y las expectativas racionales, pero si una de estas teorías hubiese sido mejor que las otras y ofrecido un flujo constante de predicciones sumamente exactas, creo que me habría dado cuenta. ¿La necesidad de un banco central independiente es una verdad establecida o sólo una moda pasajera? Desde hace décadas los economistas han propuesto

explicaciones radicalmente diferentes de los elevados índices europeos de desempleo, sin poder llegar a un acuerdo. ¿Es real que la gente trabaja menos cuando la presión impositiva es mayor? Las opiniones expertas difieren. Éstos no son asuntos situados en las fronteras de la investigación, donde siempre habrá discrepancias, sino cuestiones de larga data inscritas en el corazón de la disciplina.

El criterio 2 tampoco se cumple, como lo indica el uso del prefijo «neo», por ejemplo en neokeynesianismo o neofuncionalismo. Los criterios 3 y 4 son de fácil cumplimiento; de hecho, se cumplen en exceso. La profesionalización de la ciencia social estadounidense a la que me he referido antes tiene, en su lado negativo, una agobiante estrechez de miras y una obsesión contraproducente por el puesto que cada departamento ocupa en esta o aquella clasificación. Conjeturo que, en realidad, la ubicación en los primeros puestos de la jerarquía es un estado que, en esencia, es un subproducto de la búsqueda de conocimientos y no de posición. (Proust, como hemos visto, hacía una observación similar sobre los músicos.) Por otra parte, como el diálogo con el pasado queda relegado al subcampo de escaso prestigio de la historia del pensamiento económico o la historia apenas un poco más prestigiosa del pensamiento político, la estrechez de miras se agrava. Pero si bien las patologías del oscurantismo duro deben preferirse a las del oscurantismo blando, la elección entre unas y otras no nos corresponde.

El futuro de la ciencia social

¿Podemos hacerlo mejor? ¿Podrán las aspiraciones científicas de las ciencias sociales –la búsqueda de predicción, exactitud y precisión– cumplirse en algún momento futuro? La historia de las ciencias nos aconseja ser muy cautelosos a la hora de hacer afirmaciones acerca de lo que una ciencia puede y no puede lograr. «Todos se rieron» de Descartes cuando dijo que los animales eran máquinas, pero ¿quién ha reído el último? Cualquier intento de abordar la cuestión tendrá que ser a la vez tentativo y especulativo.

La incorporación de descubrimientos de las neurociencias permitirá sin duda a la psicología contar con un fundamento más sólido, y tal vez resuelva algunas de las controversias vigentes. Ya se ha sostenido, por ejemplo, que los escaneos cerebrales encuentran bases para el descuento cuasihiperbólico y no para el descuento hiperbólico. Quizá podríamos también establecer con firmeza la distinción entre la ira y la indignación si estuviéramos en condiciones de mostrar que estas emociones activan diferentes centros en el

cerebro. La neurofisiología de la adicción seguirá mejorando, a no dudar, nuestra comprensión de este desconcertante comportamiento autodestructivo. Adviértase, no obstante, que todas estas aplicaciones entrañan una *motivación* para la acción. Las aplicaciones a la *percepción*, como los fenómenos de completado, también son abundantes. En contraste, la neurociencia de las *creencias* proposicionales (y de las interacciones de creencias y motivaciones) no existe ni existirá en un futuro previsible. Considérese un caso trivial de actualización de creencias. Yo creía que mañana iba a llover, pero como el cielo se despejó, revisé mi pronóstico y, en consecuencia, cambié mis planes. No creo que sólo sea mi falta de imaginación (o de pericia científica) la que me hace dudar de que pronto seamos capaces de identificar siquiera la neurofisiología de este simple proceso, y debe señalarse que muchos procesos de formación de creencias son, por supuesto, enormemente más complejos. Si tengo razón, una mitad del modelo de la acción basado en las creencias y los deseos se mantendrá durante largo tiempo impermeable al enfoque neurocientífico. Lo mismo vale de manera aún más obvia si vamos más allá de dicho modelo y nos preguntamos por las «neuronas de inspiración» que generan las invenciones y las obras de arte.

Hay dos tipos de razones por las cuales las ciencias sociales no logran predecir o explicar en un sentido fuerte. Una es que aun para creencias y preferencias dadas, la acción puede seguir siendo hasta cierto punto indeterminada (es decir impredecible). En la toma de decisiones en condiciones de incertidumbre o con un alto grado de complejidad, el comportamiento puede deberse a lo que Keynes llamó nuestros «espíritus animales» y no a rasgos identificables de la situación ante la cual reaccionamos de manera identificable. En estos casos la gente actúa, claro está, sobre la base de reglas prácticas: el problema es que éstas son muchas, demasiadas. Tal vez haya, por ejemplo, puntos focales antagónicos: *haz como siempre se ha hecho* contra *donde fueres haz lo que vieres*. Decir que la gente «satisfiza» y no que «maximiza» no es decir mucho, a menos que podamos identificar *ex ante* lo que constituye un nivel satisfactorio.

Otra razón es nuestra escasa comprensión de los mecanismos de formación de preferencias. Los individuos están sujetos a propensiones antagónicas cuyas fuerzas relativas, en cualquier situación dada, suelen ser indeterminadas. Si yo te amenazo, ¿sentirás miedo o ira? Si se producen ambos efectos, ¿cuál de ellos predominará? Si me tomo un respiro en mi agitada vida profesional, ¿mi ocio será igualmente frenético o, al contrario, relajado por completo? Si mi país pasa de una dictadura a una democracia,

¿mi liberación de la autoridad política me llevará a dejar de lado también la autoridad religiosa o, por el contrario, a buscarla? He sostenido que muchas veces somos incapaces de contestar estas preguntas sobre la formación de preferencias *ex ante*, aunque a posteriori quizá podamos identificar el mecanismo dominante.

¿Podríamos reducir la indeterminación si identificáramos o bien *condiciones desencadenantes* o bien *probabilidades desencadenantes*? He examinado la primera posibilidad en el Capítulo 2, con conclusiones un tanto escépticas; permítaseme, entonces, ocuparme de la segunda. ¿Podríamos, por ejemplo, especificar una distribución de niveles satisfizadores [*satisficing*] en la población que nos permitiera al menos predecir el comportamiento agregado? Después de todo, de eso se tratan principalmente las ciencias sociales. De manera similar, si suponemos que alguna fracción de la población va a sentir miedo y no ira al ser amenazada, o que cada uno de nosotros podría reaccionar de una u otra manera, ¿no sería posible cuantificar las probabilidades?

Aun suponiendo que pudiéramos hacerlo, me parece que no ganaríamos mucha capacidad explicativa, debido a la inmensa importancia del *contexto*. Antes dije que el progreso en la ciencia es a menudo la consecuencia de una inflexible abstracción con respecto al contexto. La afirmación también es valedera para las ciencias sociales, en el sentido de que no podemos identificar propensiones o mecanismos salvo si hacemos caso omiso de muchos rasgos de la situación o creamos un ámbito en que éstos estén ausentes. La noción intuitiva de conformismo, por ejemplo, es compleja. Incluye el apoyo en los otros como fuente de información, el deseo de ser como los otros, el deseo de no asomar la cabeza y el deseo de que los otros no piensen mal de uno. Si por un propósito específico de investigación queremos definir el segundo (digamos) de estos elementos como parámetro del conformismo, tenemos que aislarlo de las otras tres dimensiones de este último y crear para ello un ámbito experimental (o encontrar uno en la vida real) donde la presencia de esas otras dimensiones sea improbable. Comprobaremos tal vez que algunas personas quieren en efecto ser como los demás, mientras que otras desean antes que nada distinguirse de ellos. Acaso podamos incluso especificar las proporciones o probabilidades. Pero en una situación de la vida real, es probable que las otras dimensiones del conformismo también estén presentes y ejerzan un efecto opuesto. Saber cómo se distribuyen las personas a lo largo de una de las cuatro dimensiones tal vez no sea de mucha ayuda para predecir un *comportamiento* conformista.

El análisis, el desmenuzamiento, es factible. La síntesis, el ensamble, es mucho más difícil, y quizás imposible.

Cuesta muy poco imaginar cómo podría responder a esta situación un estudioso inclinado a probar la posibilidad de una *ciencia* social, pero no llevaré la cuestión más adelante. Algunos nos sentimos impresionados y abrumados por la complejidad e inestabilidad del comportamiento humano. Otros tienen una creencia instintiva en una regularidad subyacente que, una vez descubierta, nos permitirá poner las ciencias sociales en un pie de igualdad con las ciencias naturales, trátese de la física, la química, la geología o la meteorología. El futuro dirá, pero he procurado dejar en claro dónde pondría mis apuestas y por qué.

Nota

* El autor se refiere a *The Coming of the Third Reich* [*La llegada del Tercer Reich: el ascenso de los nazis al poder*, Barcelona, Península, 2005], *The Third Reich in Power, 1933-1939* [*El Tercer Reich en el poder, 1933-1939*, Barcelona, Península, 2007], y *The Third Reich at War, 1939-1945*. [T.]

Proverbios y máximas citados

Antes de que te cases, mira lo que haces
Cuanto más miedo, más peligro
Cuanto más se demora la venganza, más duro es el golpe
Cuesta caro ser pobre
De padre mezquino, hijo pródigo
De tal palo tal astilla
Dios los cría y ellos se juntan
Dos ojos por uno / Ojo por ojo
El miedo es a menudo más grande que el peligro
El que cumple con su palabra en las pequeñas cosas, también lo hará en las grandes
El que miente también roba
El que miente una vez mentirá siempre
El que roba un huevo también robará una vaca
El recuerdo de peligros pasados es agradable
En todo rebaño hay una oveja negra
Equipo que gana no se toca
Hace falta una sola oveja descarriada para arruinar un rebaño
La amistad significa olvidar lo que das y recordar lo que recibes
La naturaleza propone y el hombre dispone
La necesidad es la madre del ingenio
La pasada felicidad aumenta la presente desdicha
La venganza es un plato que se sirve frío
La virtud no se conoce a sí misma
Los hombres son muy vanos, pero nada odian tanto como pasar por serlo,

Los opuestos se atraen
Los rumores suelen mentir
Nunca olvides un servicio que te hayan prestado, y olvida de inmediato el que tú has brindado
Odian también a quienes injurian
Ojos que no ven, corazón que no siente
Quien ha ofendido no puede perdonar
Recordar una desdicha es reavivarla
Si jeunesse savait, si vieillesse pouvait (Si la juventud supiera, y la vejez pudiera)
Si no está roto, no lo arregle
Si te cogen con las manos en la masa, nunca confiarán en ti
Somos muy crédulos con lo que esperamos y con lo que tememos
Una vez engañan al prudente y dos al inocente
Un mal no excedido no queda vengado
Uno ha llegado a la madurez cuando es lo bastante viejo para saber bien que no debe y lo bastante joven para, aun así, hacerlo
Vacilar es perderse
Vísteme despacio que estoy apurado

CLA•DE•MA
FILOSOFÍA

ALAIN FINKIELKRAUT	***La sabiduría del amor***
SAMUEL CABANCHIK	***Introducciones a la filosofia***
ANA MARÍA LORANDI	***Ni ley, ni rey, ni hombre virtuoso*** **Guerra y sociedad en el Virreinato del Perú. Siglos XVI y XVII**
GLADYS PALAU	***Introducción filosófica a las lógicas no clásicas***
GLADYS PALAU	***Lígicas condicionales y razonamiento de sentido común***
MARIO BUNGE	***Semántica I*** **Sentido y referencia**
MARIO BUNGE	***Semántica II*** **Interpretación y verdad**
MICHEL FOUCAULT	***La verdad y las formas jurídicas***
JEAN BAUDRILLARD	***El espejo de la producción***
ROBERT	***Historia de la idea de progreso***

Nisbet	
Norberto Bobbio	***El problema de la guerra y las vías de la paz***
Gianni Vattimo	***Introduccion a Heidegger***
Gianni Vattimo	***El fin de la modernidad*** **Nihilismo y hermenéutica en la cultura posmoderna**
Tzvetan Todorov	***Frágil felicidad*** **Un nuevo ensayo sobre Rousseau**
Jean-François Lyotard	***La posmodernidad (explicada a los niños)***
Ludwig Wittgenstein	***Sobre la certeza***
Cornelius Castoriadis	***Los dominios del hombre*** **Las encrucijadas del laberinto**
Donald Davidson	***De la verdad y de la interpretación*** **Fundamentales contribuciones a la teoría del lenguaje**
Ronald Dworkin	***El imperio de la justicia***
Jean-François Lyotard	***La diferencia***
Paul Ricoeur	***Ideología y utopía***
Jacques Derrida	***Memorias para Paul de Man***
Pier Aldo Rovatti	***Como la luz tenue*** **Metáfora y saber**

Martin Heidegger	***Introducción a la metafísica***
Gregorio Kaminsky	***Spinoza: la política de las pasiones***
José M. Beneyto	***Apocalipsis de la modernidad*** **El decisionismo político de Donoso Cortés**
Autores Varios	***Michel Foucault, filósofo***
Ian Hacking	***La domesticación del azar*** **La erosión del determinismo y el nacimiento de las ciencias del caos**
Gianni Vattimo (comp.)	***La secularización de la filosofía*** **Hermenéutica y posmodernidad**
Ernesto Mayz Vallenilla	***Fundamentos de la metatécnica***
David Gauthier	***La moral por acuerdo***
Ian Hacking	***El surgimiento de la probabilidad***
Kitaro Nishida	***Indagaciín del bien***
Hans-Georg Gadamer	***El estado oculto de la salud***
Norbert Bilbeny	***Kant y el tribunal de la conciencia***
Pascual Martínez-Freire	***La nueva filosofía de la mente***
H. R.	***El final de los grandes proyectos***

FISCHER, A. RETZER Y J. SCHWEITZER

ERNST TUGENDHAT ***Lecciones de ética***

ERNST TUGENDHAT ***Ser - Verdad - Acción***
Ensayos filosóficos

JULIO CABRERA ***Crítica de la moral afirmativa***

BERNARD SICHERE ***Historias del mal***
Con un prólogo de Julia Kristeva

MÓNICA VIRASORO ***De ironías y silencios***

MICHAEL D. RESNIK ***Elecciones***
Una introducción a la teoría de la decisión

ERNST TUGENDHAT ***Problemas***

JAAKKO HINTIKKA ***El viaje filosófico más largo***
De Aristóteles a Virginia Woolf

J. L MACKIE. ***Ética***
La invención de lo bueno y lo malo

THOMAS NAGEL ***Otras mentes***
Ensayos críticos 1964-1994

HILARY PUTNAM ***El pragmatismo***
Un debate abierto

GIANNI VATTIMO (COMP) ***Filosofía y poesía***
Dos aproximaciones a la verdad

HILARY PUTNAM ***Representación y realidad***
Un balance crítico del funcionalismo

Ernst Tugendhat	***Diálogo en Leticia***
Agnes Heller	***Una filosofía de la historia en fragmentos***
Frederic Schick	***Hacer elecciones***
Jon Elster	***La ética de las decisiones médicas***
Fina Birules (comp.)	***Hannah Arendt*** **El orgullo de pensar**
Hanna Arendt	***Tiempos presentes***
Richard Rorty	***Filosofía y futuro***
Ernst Tugendhat	***Introducción a la filosofía analítica***
Alain Badiou	***Breve tratado de ontología transitoria***
Mario Bunge	***Crisis y reconstrucción de la filosofía***
Ernst Tugendhat	***Egocentricidad y mística*** **Un estudio antropológico.**
Mario Bunge	***Emergencia y convergencia*** **Novedad cualitativa y unidad del conocimiento**
Fredric Jameson	***Una modernidad singular*** **Ensayos sobre la ontología del presente**
Immanuel Wallerstein	***Las incertidumbres del saber***
Tomás	***Contra la dominación***

Ibáñez	**Variaciones sobre la salvaje exigencia de libertad que brota del relativismo y de las consonancias entre Castoaridis, Foucault, Rorty y Serres**
J. M. Esquirol	***El respeto o la mirada atenta*** **Una ética para la era de la ciencia y la tecnología**
Alessandro Ferrara	***La fuerza del ejemplo*** **Ensayos sobre el paradigma del juicio**
Seyla Benhabib	***El ser y el otro en la etica contemporánea*** **Feminismo, comunitarismo y posmodernismo en la ética contemporánea**
Mario Bunge	***A la caza de la realidad***
Giacomo Marramao	***Kairós*** **Apología del tiempo debido**
Michel Onfray	***La comunidad filosófica*** **Manifiesto por una universidad popular**

Jon Elster es profesor en el Collège de France (Chaire de Rationalité et Sciences Sociales). Miembro de la American Academy of Arts and Sciences y miembro correspondiente de la British Academy, ha recibido, entre muchas otras, becas de la John Simon Guggenheim Foundation y la Russell Sage Foundation. El doctor Elster ha ejercido la docencia en la Universidad de Chicago y la Columbia University y ha sido profesor visitante en numerosas universidades de los Estados Unidos y Europa.

Es autor y compilador de treinta y cuatro libros; Gedisa ha publicado *El cambio tecnológico, Tuercas y tornillos, Juicios salomónicos, Lógica y sociedad, Justicia social, Psicología política, Egonomics, La democracia deliberativa, La ética de las decisiones médicas, Ulises desatado*.

Sociología
Serie CLA·DE·MA

gedisa editorial

Jon Elster
La explicación del comportamiento social
Más tuercas y tornillos para las ciencias sociales

Este libro es una versión ampliada, revisada y autocrítica de *Tuercas y tornillos: una introducción a los conceptos básicos de las ciencias sociales*, volumen del autor que conquistó la aclamación de la crítica. En veintiséis capítulos, Jon Elster describe la naturaleza de la explicación en las ciencias sociales; analiza los estados mentales –creencias, deseos y emociones- que son precursores de la acción; hace una comparación sistemática de los modelos de comportamiento basados en la elección racional con explicaciones alternativas; examina las posibles enseñanzas que las ciencias sociales pueden extraer de la neurociencia y la biología evolutiva, y revisa los mecanismos de la interacción social, desde el comportamiento estratégico hasta la toma colectiva de decisiones. Nos ofrece un panorama general de los mecanismos explicativos claves de las ciencias sociales, sobre la base de cientos de ejemplos y el recurso a una amplia variedad de fuentes: psicología, economía comportamental, biología, ciencias políticas, escritos históricos, filosofía y ficción.

En un lenguaje accesible y liberado de toda jerga, Elster aspira a la exactitud y la claridad, a la vez que elude los modelos formales. En una provocativa conclusión, defiende el carácter central de la ciencia social cualitativa en una guerra de dos frentes contra las formas blandas (literarias) y duras (matemáticas) de oscurantismo.

CC CREATIVE COMMONS

ISBN 978-84-9784-251-8
9 788497 842518
302568